前　言

进入二十世纪八十年代以来，分子生物学对现代科学的影响已遍及生物学的每一个领域，遗传学更是首当其中。1980年，美国科学家 Botstein 首次提出利用 DNA 限制性片段长度的多态性作为一种新的遗传标记来构建遗传连锁图谱。随后以美国的人类基因组计划为先导，在世界范围内兴起了生物基因组研究的热潮。分子标记技术的发展及基因组计划的实施正在对遗传学的各个领域产生重大而深远的影响。所导致的变革之一就是促使过去相互分隔的两个遗传学领域，分子遗传学和数量遗传学，发生交叉和融汇，正在发展形成一个崭新的遗传学分支学科——分子数量遗传学。本书的目的是向读者简要地介绍分子数量遗传学的形成和发展及其有关的基本原理和应用。

分子数量遗传学的基本理论既涉及包含许多数理统计原理的数量遗传学，又涉及迄今为止分子生物学领域所发展的新技术和新方法。同时，基本理论的应用又与数量性状的改良密切相关。因此，我们编写此书的目的之二是在分子遗传学、数量遗传学及遗传育种三个领域的科学工作者之间架起一座可以相互联系的纽带和桥梁，以增进不同领域科学工作者之间的理解和合作。这不仅是现代分子生物学发展的需要，也是使数量性状遗传改良迈上新台阶的基础。

尽管目前国际上还没有与本书相类似的版本供参考，但我们感到分子遗传学与数量遗传相结合并带动生物遗传育种的革命是学科发展的必然。近几年，国内许多实验室在国家高技术计划和国家水稻基因组计划的资助下，已在许多方面开展了卓有成效的研究工作。在实践中，我们深感需要有一本理论与实践相结合的综合性参考书，能够为希望涉足这一领域的研究工作者、大学生和研究生提供入门的向导。这就是我们编著本书的目的之三。

近年来，针对国际上遗传学科的发展趋势，本书作者之一徐云碧应中国科学院遗传研究所的邀请，于1992年3月至8月，举办了题为“数量遗传的分子基础”的系列讲座，第一次比较全面地综合了以分子标记为基础的数量遗传学研究方法，为这次讲座编写的讲义受到同行的欢迎，并在国内流传。期间，作者撰写了题为“发展中的遗传学分支学科——分子数量遗传学”的论文，被中国遗传学会召开的全国青年遗传学家学术研讨会录用，并收入《遗传学基础理论问题讨论文集》（北京师范大学出版社，1993年12月）。同年10月，作者在浙江农业大学为选修数量遗传学的硕士研究生开设了“数量遗传的分子基础”专题讲座，比较系统地讲授了分子数量遗传学的基本内容，这一专题讲座后来被作为研究生数量遗传学课程的主要内容确定下来。1994年10月，作者在浙江农业大学生物数学中心与中国农学会生物数学学会在杭州联合举办的生物数学专题研讨会上，比较系统地介绍了分子数量遗传学的分析方法及其研究进展。并于同年11月分别在中国遗传学会于上海农学院召开的全国中青年教师遗传学进展报告会及中国作物学会于南京农业大学召开的第三届全国青年作物遗传育种学术会议上，作了题为“分子数量遗传学研究进展与展望”和“分子标记在植物数量性状遗传改良中的应用”的报告。本书正是在上述努力的基础上合作完成的。

在本书完稿之际，我们要感谢国家高技术计划和水稻基因组计划多年来给予我们研究工作的资助，使我们有机会在与本书有关的领域开展研究，并积累了大量的实验结果，使本书有可能利用大量的第一手材料作为实例，并把我们的研究体会奉献给读者。

《分子数量遗传学》作为第一次尝试的产物，我们感到只是尽了我们的努力。需要说明的是，由于我们所从事的研究领域及专业所限，我们主要以植物为例说明有关的基本理论和方法。同时，书中所提供的应用实例，则主要以我们所研究的对象——水稻为主。从广义上讲，书中所涉及的基本理论和方法，既适合于植物，也适合于或者可以推广到包括动物在内的其它生物。

我们感谢中国科学院院士、中国农业科学院庄巧生研究员在我们写作过程中所给予的关心和帮助，感谢国家高技术计划生物技术领域青年科学基金、国家高技术计划和基因组计划对研究工作及本书出版的资助。同时，我们恳请各位读者和同行对本书提出批评和建议。

作　者

一九九四年十二月六日

（京）新登字060号

内 容 提 要

本书是当代遗传科学的最新进展与作者科学研究成果的综合产物。作者回顾和展望了分子遗传学与数量遗传学相互融合和发展及分子数量遗传学的形成；论述了现代分子生物学所发展的各类分子标记及其在遗传图谱构建中的作用；系统描述了利用遗传标记构建连锁遗传图谱的统计原理、方法和计算机软件，几类可以用于遗传作图的永久性群体及其特性；详细阐述了 QTL 作图的统计理论、方法、软件开发及应用实例，包括：平均值差异法，方差分析法， 以单一标记为基础的最大似然法，以侧连标记为基础的区间作图法，复合 QTL作图等。 在分子数量遗传学理论的实际应用方面，探讨了分子标记在植物育种中的应用，基因分离克隆的基本方法，以分子标记图谱为基础的分子克隆及其对未来作物改良的可能影响，数量性状的基因克隆等问题。 附录提供了 Quick Basic 语言编写的有关统计分析的计算机程序、练习题及详细的参考文献。本书可作为高等学校生物类高年级学生、研究生的课程教材，也是分子生物学、分子遗传学、数量遗传学以及遗传育种工作者的重要参考书。

目　　录

第 1 章　发展中的遗传学分支学科 —— 分子数量遗传学 (1)

1.1　遗传学领域两门分支学科的发展与分化 (1)
1.1.1　数量遗传学的建立和发展 (1)
1.1.2　分子遗传学的发展 (3)
1.2　遗传标记的发展和分子数量遗传学 (5)
1.3　现阶段分子数量遗传学研究的主要内容和方法 (6)
1.3.1　遗传标记与数量性状的关联及其检测 (7)
1.3.2　QTL 作图 (7)
1.3.3　QTL 克隆 (7)
1.3.4　QTL 序列分析 (7)
1.4　分子数量遗传学的发展前景与展望 (8)

第 2 章　数量性状的生统遗传学基础 (10)

3.1　数量性状及其遗传特点 (10)
2.1.1　数量性状与数量遗传学 (10)
2.1.2　群体及其遗传参数 (10)
2.2　群体平均数及其遗传分量 (13)
2.2.1　平均数的计算公式 (13)
2.2.2　群体表型值的分解 (13)
2.2.3　基因效应 (14)
2.2.4　近交群体世代平均数的遗传分量 (15)
2.3　群体方差及其遗传分量 (16)
2.3.1　方差的概念及其计算 (16)
2.3.2　群体性质和方差分量 (18)
2.3.3　近交群体的方差分量 (18)
2.3.4　群体性质和协方差分量 (20)
2.4　数量遗传模型及其检验 (22)
2.4.1　加性-显性遗传模型 (22)
2.4.2　加性-显性-上位性模型 (23)
2.5　遗传变异的传递与选择 (24)
2.5.1　有效因子数或最小基因数 (24)
2.5.2　显性度 (25)

2.5.3 遗传率及其估计 (25)
2.5.4 直接选择与选择响应 (26)
2.5.5 遗传相关与间接选择 (29)
2.5.6 选择指数与综合选择 (30)
2.6 基因型与环境互作 (31)
2.6.1 基因型与环境互作的表现 (31)
2.6.2 基因型与环境互作的遗传模型 (32)
2.6.3 基因型与环境互作效应的估算 (33)
2.6.4 基因型与环境互作分析在植物育种中的应用 (34)

第 3 章 遗传标记 (35)

3.1 形态标记 (36)
3.2 细胞学标记 (37)
3.3 生化（同工酶）标记 (41)
3.4 分子标记 (41)
3.4.1 RFLP 标记 (43)
3.4.2 AFLP 标记 (44)
3.4.3 RAPD 标记 (48)
3.4.4 重复序列 (51)
3.4.5 STS (52)
3.5 各类遗传标记的特点及其利用 (52)

第 4 章 遗传图谱构建 (55)

4.1 作图群体的建立 (55)
4.1.1 亲本的选择 (55)
4.1.2 组合的配制 (55)
4.1.3 作图群体的类型 (56)
4.2 连锁作图的理论基础 (57)
4.2.1 Sutton-Boveri 染色体遗传理论 (57)
4.2.2 基因重组和连锁 (58)
4.2.3 影响交换和重组的因素 (59)
4.2.4 遗传图距的相对性 (61)
4.3 遗传图谱的制作 (61)
4.3.1 分子标记分离数据的收集和处理 (61)
4.3.2 连锁的两点测验 (63)
4.3.3 利用 MLE 估计重组值 (66)
4.3.4 似然比与连锁检验 (70)
4.3.5 多点分析与基因直线排序 (71)

4.3.6 交换干扰和作图函数 (73)
4.3.7 染色体着丝粒和端粒的定位 (75)
4.3.8 分子标记连锁群的染色体定位 (77)
4.3.9 经典遗传图谱与分子遗传图谱的综合 (77)
4.4 用于构建遗传图谱的计算机程序包 (78)
4.4.1 人类 CEPH 型数据的连锁作图 (78)
4.4.2 F_2 数据的连锁作图 (79)
4.5 与遗传图谱构建有关的几个问题 (81)
4.5.1 构建饱和图谱所需基因座位数 (81)
4.5.2 群体或组合数目 (82)
4.5.3 群体大小 (82)

第 5 章 群体的遗传结构 (86)

5.1 群体性质及其分类 (86)
5.1.1 根据群体遗传组成划分 (86)
5.1.2 根据群体的持续性划分 (86)
5.1.3 根据群体遗传背景划分 (87)
5.1.4 根据群体来源划分 (87)
5.2 DH 群体 (89)
5.2.1 DH 系的产生途径 (89)
5.2.2 DH 系的评价 (89)
5.2.3 DH 群体的数量遗传学 (90)
5.2.4 DH 群体在分子作图中的应用 (90)
5.3 RIL (SSD) 群体 (92)
5.3.1 近交及其遗传效应 (92)
5.3.2 RIL 的产生途径 (96)
5.3.3 RIL 群体中图距与重组率的关系 (96)
5.3.4 利用 RIL 构建遗传图谱 (97)
5.4 NIL 群体 (98)
5.4.1 回交及其遗传效应 (98)
5.4.2 NIL 作图策略 (100)
5.4.3 NIL 作图的理论问题 (100)
5.4.4 应用 (105)
5.4.5 产生 NIL 的其它途径 (107)
5.5 几类群体的连锁作图及其效率 (108)
5.6 群体处理方法 (110)
5.6.1 选择性基因型分析 (110)
5.6.2 分离群体分群分析法 (111)

第 6 章 QTL 作图: 均值-方差分析法 (114)

6.1 基本假定 (115)
6.2 标记-性状关联的检测 (116)
6.2.1 回归分析法 (116)
6.2.2 单向分组方差分析法 (118)
6.2.3 均值和方差比较法 (119)
6.3 重组率及 QTL 遗传参数的估计 (121)
6.3.1 利用均值估计 (122)
6.3.2 利用方差估计 (128)
6.3.3 连锁检测所需的最小植株数 (123)
6.4 利用标记频率检验连锁 (125)
6.4.1 基本原理 (125)
6.4.2 试验设计 (125)
6.5 应用与实例 (126)
6.5.1 应用概况 (126)
6.5.2 数据实例 (129)

第 7 章 QTL 作图: 矩法和最大似然法 (138)

7.1 原理与统计方法 (138)
7.1.1 矩估计法 (139)
7.1.2 最大似然法 (140)
7.1.3 似然比检验 (142)
7.1.4 QTL 基因效应的估计 (143)
7.1.5 程序设计 (144)
7.2 Monte Carlo 模拟 (144)
7.3 应用与实例分析 (147)
7.3.1 应用 (147)
7.3.2 实例分析 (147)

第 8 章 QTL 作图: 区间作图法 (158)

8.1 遗传学原理与统计方法 (158)
8.1.1 基本假定 (158)
8.1.2 似然函数 (159)

8.1.3 LOD 检验 (161)
8.2 区间作图软件 (162)
8.2.1 QTLMAP-I (162)
8.2.2 Mapmaker/QTL (162)
8.3 应用 (165)
8.3.1 应用 (165)
8.3.2 实例 (167)

第 9 章 QTL 作图: 复合区间作图法 (171)

9.1 区分多个连锁 QTL 的理论基础 (172)
9.1.1 统计模型 (172)
9.1.2 偏回归系数 (173)
9.1.3 偏回归系数的抽样方差 (175)
9.1.4 偏回归系数的抽样相关 (177)
9.1.5 多元回归的性质 (178)
9.1.6 回交群体以外的其它群体 (179)
9.2 复合区间作图 (179)
9.2.1 统计原理 (179)
9.2.2 零假设下检验统计量的特性 (181)
9.3 模拟分析 (185)
9.3.1 方法 (185)
9.3.2 结果 (187)

第 10 章 植物数量性状的选择与育种 (190)

10.1 回交导入有利基因 (191)
10.1.1 标记辅助技术的特点 (191)
10.1.2 所需回交代数 (192)
10.1.3 回交育种计划的设计 (194)
10.2 近交选择 (195)
10.2.1 单标记选择 (195)
10.2.2 多标记选择 (195)
10.2.3 多个性状选择 (197)
10.2.4 寡基因抗性的选择 (197)
10.3 图式基因型和全基因组选择 (199)
10.3.1 推测图式基因型的基本前提 (199)
10.3.2 推测图式基因型的基本假定 (200)
10.3.3 图式基因型实例 (201)
10.3.4 利用图式基因型进行全基因组选择 (202)

10.4 选择效果及其评价 (203)
10.4.1 玉米群体数量性状的选择响应 (205)
10.4.2 长期选择结果的评价 (205)
10.5 杂种优势分析 (205)
10.5.1 杂种优势的分析 (205)
10.5.2 杂种优势的预测 (206)
10.5.3 与杂种优势有关的 QTL (207)
10.6 无性繁殖品种的改良 (207)
10.7 生物起源与进化的研究 (207)
10.8 其它应用 (208)
10.8.1 亲本控制和品种鉴别 (208)
10.8.2 病原菌鉴别 (209)
10.8.3 利用外源种质的新途径 (209)
10.8.4 育种实践中的其它问题 (210)
10.8.5 植物群体生物学 (210)
10.8.6 无性系变异 (210)
10.8.7 成本效益分析 (210)

第 11 章 QTL 分离克隆的战略 (212)

11.1 植物基因分离的一般方法 (212)
11.1.1 以基因功能为基础的克隆方法 (212)
11.1.2 以 mRNA 或蛋白质表达水平为基础的克隆方法 (213)
11.1.3 以 DNA 插入为基础的克隆方法 (213)
11.1.4 减法杂交 (215)
11.2 以图谱为基础的克隆方法 (216)
11.2.1 利用紧密连锁的分子标记跟踪基因 (216)
11.2.2 利用脉冲凝胶电泳进行大尺度物理作图 (218)
11.2.3 酵母人工染色体大 DNA 片段的克隆 (219)
11.2.4 大片段 DNA 分子的转移 (220)
11.2.5 实例 (220)
11.3 随机 cDNA 序列分析的应用 (222)
11.4 QTL 效应分析与歧化选择 (223)
11.4.1 理论 (224)
11.4.2 实例 (225)
11.5 QTL 克隆的可能途径 (228)
11.5.1 以图谱为基础的 QTL 克隆 (228)
11.5.2 以随机 cDNA 序列分析为基础的 QTL 克隆 (228)
11.5.3 利用插入突变克隆 QTL (228)
11.5.4 其它方法 (229)

11.6 与 QTL 克隆有关的基础研究 (230)

附录

A1 QTL 作图软件 (233)
A1.1 建立数据文件 (DATASET.BAS) (233)
A1.2 均值比较 (QTLMEAN.BAS) (240)
A1.3 单因素方差分析 (QTLVAR.BAS) (245)
A1.4 矩估计法 (QTLMS.BAS) (247)
A1.5 最大似然法 (QTLMLE.BAS) (253)
A1.6 区间作图法 (QTLINT.BAS) (260)
A2 进一步阅读的建议 (267)
A3 练习题 (269)

第 1 章　发展中的遗传学分支学科
—— 分子数量遗传学

十九世纪下半叶，孟德尔(G. Mendel)以豌豆为材料进行杂交试验，以质量性状为研究对象，发现了遗传因子的传递规律，为遗传学的产生奠定了基础。一个多世纪以来，遗传学的发展围绕着由单个或少数基因控制的质量性状的遗传现象由浅入深、由表及里，经历了细胞遗传学、微生物遗传学和分子遗传学三个重要的发展时期，成为包括众多分支学科在内的重要生物学科。而长期以来以数量性状为研究对象的数量遗传学，几乎完全独立于遗传学发展的主流之外，形成了一个独立发展的遗传学分支。直至本世纪八十年代初，分子遗传学开始向高等动植物的基因组进军，数量遗传学和分子遗传学这两大分支才开始相互渗透和融汇。本章将讨论这两门分支学科的发展、分化；分子数量遗传学的形成；分子数量遗传学研究的主要内容和方法；分子数量遗传学的发展前景和展望。

1.1　遗传学领域两门分支学科的发展与分化

分子数量遗传学是在数量遗传学和分子遗传学的基础上发展起来的一门崭新的遗传学分支学科。为了对这门学科的形成和发展有一个清晰的轮廓，我们首先回顾一下这两门学科各自的发展史，然后讨论遗传标记的发展对这两门学科相互交叉和融汇的影响。

1.1.1　数量遗传学的建立和发展

从各个遗传学分支学科发展形成的时间来看，数量遗传学是最为古老的一个分支，甚至可以说它的萌芽早于遗传学科本身。在孟德尔以前的十八世纪，许多植物育种家为改良植物进行了大量的杂交试验，在其杂种后代中就观察到了大量数量性状的连续变异。后来F. Galton进行人类身高的研究，不过从这一数量性状的分析未能奠定遗传学大厦的基石。但他发现了数量性状至少是可以部分遗传的。他的有关亲子遗传的定律以及所创造的回归与相关的工具成为后来认识和分析数量性状的宝贵财富。后来众多遗传学的先辈们经过几十年的努力，才将孟德尔原理与数理统计相结合，形成了以数量性状为研究对象的遗传学分支学科——数量遗传学。由于数量性状一般由多基因控制，单个基因的效应较小，易受环境影响，无法个别地区分基因的作用，长期以来数量遗传学家将这多个基因作为一个整体进行研究，并采用反映群体特征的平均数、方差和协方差等来描述多个基因的遗传特征。

数量性状和质量性状的遗传研究在起始时间上几乎没有什么先后之分。然而早在1865年孟德尔就根据质量性状的研究结果提出了分离和独立分配规律，而对于数量性状直到1909年才由H. Nilsson-Ehle 提出了解释其遗传的多基因假说。这是由于数量性状的遗传远较质量性状复杂，加上当时由于没有把数学原理和统计理论引入杂交试验资料的分析，故无法对数量性状的遗传作出解释。值得指出的是，正是由于支配质量性状的孟德尔定律

的重新发现，并以此为基础，引进生物统计方法，才使数量性状的遗传规律得以逐步发现。

F. Galton 第一次试图用数学方法去处理群体中的遗传学问题，他在生物统计学领域的基本研究工作对数量遗传学具有深远的影响。1889年，Galton把群体的定量测量法引进了遗传学，并把群体当成一个单位来处理，建立了一系列定律。其中最重要的是回规律(Law of regression)。二十世纪初，在遗传和进化问题上，形成了以W.Bateson为首的孟德尔学派和以 K. Pearson为首的Galton学派，双方论争的焦点集中在前者认为不连续变异是进化的重要因素，并认为连续变异是不能遗传的；后者认为连续变异是进化的重要因素，而孟德尔遗传法则对于连续变异不适用。这场争论一直持续到1909年由W.Johannsen和 H. Nilsson-Ehle 分别提出纯系学说 (Pure line theory) 和多基因假说 (Polygene hypothesis)后才告结束。W. Johannsen从十九个菜豆(*Phaseolus vulgaris*)纯系粒重的研究得出结论，任何一个性状的表现型都是基因型与环境共同作用的结果；基因的作用是可以遗传的，在混杂的群体中选择有效，环境的作用是不能遗传的，在纯系内选择无效。基因型的不连续性可以由环境的作用体现在表现型上的连续性。Nilsson-Ehle (1909)、East(1915)、Emerson和East (1913)、Sax(1923)等分别以小麦粒色、烟草花冠长度、玉米果穗长度的遗传及对菜豆种子大小与粒色连锁的遗传研究表明，数量性状是受微效相似、彼此独立、无显隐性的多基因支配，这些多基因存在于染色体上，也服从于孟德尔的遗传法则。

1900年以后，R. A. Fisher及其学生为数理统计方法在许多学科特别是农学、遗传学和生物学中的应用作出了贡献。1918年，R. A. Fisher 把生物统计学与遗传学结合起来，研究人类亲属间的相关问题，进一步从理论上证实了多基因学说的重要性。并对数量性状的连续变异进行数学分析，把遗传方差(Genetic variance)分解为加性方差、显性方差和上位性方差三个部分。1925年R. A. Fisher的《研究工作者统计方法》一书为数量遗传的研究提供了基本的统计方法。

S. Wright第一个把通径系数(Path coefficiency)分析的方法应用于实际试验资料，在1921年和1936年分别发表了“交配系统”和“孟德尔群体的进化”等重要论文。他通过定义近交系数——结合配子的相关系数，解释了近交对杂合性的影响。

到五十年代，按照East 和Hayes(1912)的倡导，在玉米育种中已广泛采用异交和近交，并且Comstock和Robinson在杂交玉米和异交群体中进行了比较正规的数量遗传分析。Lush及其同事以及Lerner等采用数量遗传的方法来分析农畜重要的经济性状，并倡导了选择程序。根据Fisher等(1932)的工作，K. Mather发展了关于植物、动物杂交资料的分析方法，并在爱丁堡开始数量遗传应用于动物育种的研究。他把孟德尔方法扩展到数量性状的遗传分析，其主要方法是进行稳定遗传品系之间的杂交，从而发现起作用的基因特性，他提出了“多基因体系”(Polygenic system)的假说，发展了多基因假说。K. Mather强调连锁的重要性，他假定遗传相关只是由于连锁而不是由于多效性，并且为解决分析连锁基因的难题引入了有效因子 (Effective factor) 的概念。Mather(1949)、Kempthorne (1957)和 Hayman(1958)、Gamble (1962)等分别提出了加性、显性和上位性的遗传模型及其分析方法。

到这时，Lush (1937,1945) 的《动物育种设计》，Lerner(1950)的《群体遗传学与

动物改良》，Mather (1949)的《生统遗传学》，Li (1955) 的群体遗传学，Kempthorne (1957) 的《遗传统计学》，以及 Falconer (1960) 的《数量遗传学》概括了这一时期数量遗传学的主要成果，奠定了数量遗传学的理论基础。

将数量遗传学应用于动物育种的主要开拓者是J. L. Lush，他于30年代开始这方面的研究。他关于乳牛遗传率(Heritability) 估计的论文是原理、方法、结果与应用的极好结合。他创造了遗传率这一术语，说明了方差的分解，遗传率的含义及遗传率估计的公式。将数量遗传学方法扩展到植物群体的主要学派包括北卡罗来纳州立大学的R. E. Comstock和H. F. Robinson所领导的，以及伯明翰大学的K. Mather和J. L. Jinks所领导的研究团体。前者以玉米为主要材料，推导出从随机交配群体及F_1杂交产生的群体分解遗传方差的公式。Griffing (1956)对双列杂交设计作了综合性的全面整理和研究，提出了四种试验方法及其相应的几种分析方法。Hazel (1943)在关于选择指数中评述的性状间的相关，包括了表型相关的和基因型相关的概念。 Comstock 和Moll (1963)等关于基因型×环境互作(Genotype×environment interaction) 的研究，将效应表述为表型与基因型值离差的组成部分。K. Mather和J. L. Jinks提出了一整套分析杂种优势的生统遗传学理论。

七十年代以来，随着概率论、线性代数、多元统计和随机过程等的逐步应用，使数量遗传学又有了很大发展，而电子计算机的广泛使用使过去无法用人工进行的运算和分析成为可能。 Kearsey和Jinks (1968)所倡导的三重测交法(Triple test cross)是NCⅢ的一种扩展，被用于上位性、基因型×环境互作以及杂种优势等的分析。Lin和Binns(1986)和Westcott(1986)评价了过去广泛采用的基因型×环境互作分析方法，并倡导聚类分析等多元分析(Multivariate analysis)方法。 1987年6月召开的国际数量遗传会议更是反映了数量遗传研究从单变元向多变元发展的趋势。近年来，混合模型 (Mixed model) 和复合模型(Mixture model)等复杂统计模型在数量遗传研究中的应用，正在发展包括三倍体胚乳性状、主效基因和多基因控制的复合性状(或称为质量-数量性状)等的复杂遗传模型及其分析方法。然而遗憾的是，复杂的数学方法和手段终究不能使抽象的数量基因具体化、物化到分子水平的DNA序列。毫无疑问，数量遗传学在这一方面的升华需要新的推动，才能步出困境。

1.1.2 分子遗传学的发展

与数量遗传学的研究迥然不同，在微生物遗传学基础上发展起来的分子遗传学，则是根据物理、化学的原理来解释遗传现象，并在分子水平上研究遗传机制及遗传物质对代谢过程的调控。它认为基因是一种DNA片段，因此分子遗传学是在生命信息大分子的结构、功能及相互关系的基础上研究遗传与变异的科学。 实质上， 分子遗传学所研究的就是基因的本质(包括基因的化学性质、结构和组织)、基因的功能以及基因的变化等问题。

孟德尔在豌豆杂交试验中选择成对的性状，在杂交后代中对不同类型的个体进行统计分析，由此而阐明了决定不同性状的遗传因子是独立遗传的，具有颗粒性。无论是数量遗传学还是分子遗传学均能在孟德尔简单的杂交试验中找到自己的渊源。前者发端于对连续变异的数量性状的数理统计分析，后者则是对遗传因子的本质和作用机制深入研究的结果。分子遗传学与经典遗传学是一脉相成的，在1900年孟德尔定律再发现后不久， 遗传学

(Genetics)和基因(Gene)两个新名词应运而生，从此遗传学与基因结下了不解之缘，遗传学的研究几乎完全是围绕着基因的传递、基因的化学本质、基因的表达而展开的。摩尔根(T. H. Morgan)及其同事利用果蝇作为研究材料，通过对突变基因的遗传分析和细胞遗传学的手段，将抽象的遗传单位——基因落实到染色体上。他们揭示了在基因传递过程中连锁和交换的遗传现象，并在此基础上建立了果蝇染色体的基因连锁图。由摩尔根等提出的遗传学的新概念和新方法成为经典遗传学的核心，并发展成为一个完整的知识体系。利用这个体系，可以解释各种生物常见的遗传现象，也包括数量性状的遗传。然而在这一体系中，仍然留下一片空白有待探索，即在经典遗传学中位于染色体上的基因只是一个形式上不可分割的单位。直到本世纪四十年代，遗传学仍然没能回答基因是什么，基因是如何作用的？而正是对这两个问题的探索启动了分子遗传学的发展。

从三十年代开始，就有一批物理学家出于一种科学的使命感对生命的本质产生兴趣，希望阐明基因的物理和化学本质。他们的跨学科转行，给生命科学的发展带来了一场革命。分子遗传学正是在这一转机中脱颖而出。 1944年，O. T. Avery在肺炎双球菌的转化实验中证明了脱氧核糖核酸(DNA)是导致转化的遗传物质。1953年，J.D. Watson和 F.H. C. Crick 提出了DNA分子结构的双螺旋模型，通常DNA双螺旋模型的提出被作为分子遗传学的真正开端。1955年，S. Benzer对大肠杆菌T4噬菌体的基因重组分析，发现基因虽然是功能的单位(顺反子)，但不是突变的单位(突变子)和重组的单位(重组子)，他在分析基因内部的重组子时已达到DNA链上三个核苷酸的水平。与此同时，对突变的分析也达到了组成DNA的核苷酸碱基的水平。

几乎在同一时刻，有关基因作用的研究也取得了突破性的进展。G. W. Beadle和E. L. Tatum 通过对粗糙链孢霉营养缺陷型的研究，在四十年代初提出了一个基因一个酶的假设，这一假设沟通了生物化学中蛋白质合成的研究与遗传学中基因功能的研究。根据这一假设，可以提出两个问题。第一，在蛋白质多肽链中氨基酸排列顺序的信息是如何储存在DNA分子中的？第二，在DAN分子上的遗传信息是如何通过蛋白质合成转变为氨基酸顺序的？对前一问题的回答导致遗传密码的解码，对后一问题的求解则阐明了在细胞内大分子之间信息传递的过程——转录和翻译。 这为信息流从DNA→RNA→蛋白质的中心法则奠定了基础。这也是分子遗传学在五十年代后期和六十年代初期的两个最为活跃的研究领域。1961年，J. Monod和F. Jacob 根据在大肠杆菌和 λ 噬菌体中的研究结果提出了乳糖操纵子模型，从理论上解决了基因表达调控的难题。因此到1965年纪念孟德尔的“植物杂交实验”的论文发表一百周年时，分子遗传学已经取得了辉煌的成果，成为生物学中最为活跃的前沿学科。 从这一历史不难看出，分子遗传学是微生物遗传学和生物化学互相结合共同发展的结果。

到了七十年代初，分子遗传学又迎来了一个技术革命的新时期，到1993年，一个崭新的技术体系形成了。它的核心是离体重组DNA (Recombinant DNA) 的技术。利用这项技术可以人为地转移和改造基因从而达到在遗传上改造生物的目的，因此又称为遗传工程。遗传工程一方面成为生物高科技产业的支柱，另一方面又使分子遗传学进入一个全新的发展时期。从七十年代后期到八十年代，分子遗传学的研究突破了微生物的范畴，在人类、动物和植物等各种不同的真核生物的类型中全面展开，使真核生物的分子遗传学成为分子生

物学的热点，形成了新的前沿。由于利用新的技术手段研究新的材料，在遗传学中新的发现又接踵而来．首先是对基因本质的认识进一步深化。1977年，在噬菌体中发现了重叠的基因，一个基因的序列可被包含在另一个基因中；两个基因的序列可能部分跨叠。同年在哺乳动物中又发现基因的编码序列不是连续的，许多真核生物的基因可分解为编码的外显子和不编码的内含子。在这一时期，还在分子水平上证实了由B. McClintock在玉米中发现并推论的转座因子。分子遗传学的研究还渗透到生物学的其他学科中，使得免疫学、发育生物学、细胞生物学和神经生物学等各个学科都提高到一个新的发展层次。如在免疫学中对抗体基因的结构和功能的阐明，在发育学研究中异位同形基因 (Homeotic genes) 和异位同形盒(Homeobox) 的发现及其功能的分析，都是与分子遗传学进展密切相关的。

当我们仔细分析进入九十年代的分子遗传学的发展时，可以发现，分子遗传学已经融化到生物学的各个学科中了，几乎成为其他学科的一个内在的组成部分。然而对遗传本质的理解，遗传学还留下一大一小两个问题，有待从分子水平上进行全面的研究。大问题就是基因组的奥秘。美国在1990年开始的人类基因组计划，首先吹响了向基因组全面进军的号角。另一个较小的问题就是数量遗传学。长期以来，分子遗传学把数量性状撇在一边，视而不见。主要的困难在于：直到八十年代中期，分子遗传学还无法对数量性状的基因进行定性的研究，因为数量性状的基因仍然是难以捉摸的抽象单位。然而从分子遗传学的角度阐明数量性状遗传的本质，是分子遗传学在今天面临的一大挑战，分子遗传学终究不能也不应该回避数量性状遗传这个老大难的问题，这正是本书要讨论的主题——分子数量遗传学。

综上所述，由于数量遗传学和分子遗传学两个学科在研究对象、研究手段和方法等方面存在着天差地别，几十年来遗传学的发展使这两个遗传学分支学科各自沿着自己的方向发展，使这两个分支学科的同行被此疏如陌生人，缺乏相互理解和沟通。不仅如此，由于近代数量遗传学的发展吸收了线性代数、概率论、多元统计、随机过程等大量复杂的数理统计理论，并借助于电子计算机进行日益复杂的试验设计和统计分析，使这门学科的发展更加难以为其他的遗传学家所接受，从而使其与其他学科的关系逐渐疏远。在遗传学蓬勃发展的今天，许多有识之士呼吁要加强与人类密切相关的重要数量性状的遗传研究，期望着将现代分子遗传学发展的包括分子克隆在内的各种技术应用于数量性状的遗传操纵，使遗传学为人类作出更大的贡献。

1.2 遗传标记的发展和分子数量遗传学

传统的数量遗传学采用数理统计学的方法把控制某一性状的多个基因作为一个整体进行研究。这种方法无法鉴别出单个数量基因以及与之有关的染色体片段，更难确定数量性状基因座位 (Quantitative Trait Locus, QTL) 在染色体上的位置及与其他基因的关系，因而也无法采用现代分子遗传学所发展的基因分离、克隆等方法对数量性状进行遗传操纵(徐云碧,1990)。解决这一问题的突破口在于能否将影响数量性状的多个基因剖分开来，将其定位在染色体上，并最终确定与之有关的DNA序列。借助于已知染色体座位的遗传标记去追踪每一染色体片段的传递，不仅可以对QTL所在的基因组进行一段一段地分析，而

且可以直接测量过去所不能鉴别的各染色体区段的效应，从而确定其在染色体上的位置、单个效应及互作效应，这就向QTL分子图谱的制作以及进而为物理图谱的制作和数量性状的遗传操纵迈开了第一步。事实上，数量遗传学家为实现上述目标至始至终在进行着不断的努力。以经典的形态学标记和细胞学标记为材料，在八十年代以前就曾对QTL进行过一些定位研究(Law1966;Thoday1961)，但由于自然群体中这些标记十分稀少以及其他缺陷，未能在生物界得到广泛应用。随着生化遗传学的发展，同工酶作为一种分子标记曾应用于番茄(Tanksley等1982)和玉米(Stuber等 1980)等的QTL遗传分析。近十年来,基因克隆技术使许多生物学领域获得了重大进展，其中最激动人心的是利用染色体DNA序列本身的多态性作为遗传标记进行连锁作图。这些标记包括DNA限制性片段长度多态性(Restriction Fragment Length Polymorphism, RFLP; Botstein等1980)、随机扩增多态性DNA (Random Amplified Polymorphic DNA, RAPD; Williams等1990,Welsh和McClelland 1990)、串联重复序列(Tandem Repeats Sequence, TRS; Weber 1990a)或称简单重复序列（或微卫星序列)、单链构象多态性(Single Strand Conformation Polymorphism, SSCP; Orita等1989) 等。它们具有存在于任何发育阶段、组织及器官，没有有害的表型效应，不受外界环境干扰，等位基因一般为共显性等特点。以RFLP为标记，目前已完成了包括玉米、番茄、水稻、大麦、油菜等在内的二十余种主要作物的分子遗传图谱。完整和饱和分子图谱的建立，已有可能将多个数量基因剖分为离散的孟德尔因子，并借助于现代分子遗传学的方法对其进行遗传操作，从而使遗传学历史上这两个相互分隔的领域融为一体。

随着不同植物RFLP图谱的建立，分子标记在QTL遗传研究中的应用也日渐增多，给一度处于瓶颈的数量遗传学研究注入了活力。许多分子遗传学家、生物统计学家以及经典的数量遗传学家纷纷在这一领域开展研究。几年来在国际上已形成了一股研究QTL的热潮。在很短的时间内，利用分子图谱分析QTL的各种理论、模型和方法纷纷出笼。形势的发展促使过去十分陌生的两门遗传分支的科学家走到一起，探讨他们所共同关心的问题。从1988年11月在冷泉港实验室召开的第一次有分子遗传学、育种学和数量遗传学三方面专家参加的“植物遗传学问题中分子标记的发展与应用”研讨会开始，迄今已举办了多次生物技术与数量遗传学的小型国际学术研讨会，并在 1993年8月召开的第17届国际遗传学大会上，利用分子标记研究数量性状的遗传成为会议的主要议题之一。由此可见，将分子遗传学所发展的方法和技术应用于数量性状的遗传操纵，正在日益受到遗传学家和育种学家的重视。可以预期不久的将来，数理统计和现代生物技术综合应用于数量遗传操纵的分子数量遗传学将发展成为一门崭新的遗传学分支学科。

1.3 现阶段分子数量遗传学研究的主要内容和方法

可以预期，分子数量遗传学研究的主要内容和方法将随着遗传学的发展而变化。但在未来十年内其研究将主要围绕遗传标记与数量性状的相互关联，数量性状基因(QTL)的定位(作图)，QTL的克隆，QTL序列分析以及与此有关的理论和实验技术体系的建立，分子数量遗传理论在生物遗传改良中的应用等。这些研究在很大程度上受制于生物全基因组研究

的进展，特别是家鼠、拟南芥等重要模型生物及人类等基因组研究的进展。

1.3.1 遗传标记与数量性状的关联及其检测

为了简化数量性状的遗传分析，提高遗传改良的效率，最简单而有效的途径之一是建立遗传标记与数量性状之间的相互关联关系。这种相互关联分析可以利用单个或少数标记进行，也可以利用大量遗传标记所构建的完整连锁遗传图谱。为此，首先要由目标性状有差异的品系产生分离群体；然后对群体中各个体或品系进行该性状及在基因组上分布的遗传标记进行考查记载；最后根据标记基因型的种类或目标性状的表现将分离群体分成几个亚群体，采用方差分析或t测验比较不同标记基因型间数量性状均值或高低表现型的亚群体间标记基因频率的差异，以寻找性状与一个或多个标记遗传模式之间的相关。目前所采用的方法主要是以标记为基础(Marker-based)的MB法和以性状为基础(Trait-based)的TB法。一旦发现了标记与数量性状之间的显著相关，就可以将这种关联应用于标记辅助的选择之中，其应用效果取决于标记-性状关联的密切程度。有关内容详见第6章。

1.3.2 QTL作图

QTL作图就是采用类似于质量基因定位的方法将QTL定位在遗传图谱上，确定QTL与遗传标记间的距离(以重组率或图距表示)。采用适当的统计方法，利用两个品系间分离群体的资料可以定位 QTL与标记座位之间的重组率，并确定QTL不同等位基因遗传参数的估计值。根据在统计分析中同时采用的标记数目的不同，可以将QTL作图的方法分为以单一标记为基础的方法、以两个相邻标记为基础的方法以及以多个标记为基础的方法。根据所采用的统计分析方法的不同可将有关方法分为方差分析法、回归分析法和最大似然法等。根据分析所包含的标记区间数可分为零区间作图，单区间作图和多区间作图。目前应用最多的是以两个相邻标记为基础，采用最大似然法进行的单区间作图。此外，为提高QTL作图效率，还发展了将不同方法结合起来的综合分析方法，如将以两个标记为基础的区间作图与多元回归方法结合形成的QTL复合作图法等。有关内容详见第7～9章及最近的两篇综述(徐云碧 1994c；徐云碧 1995)。

1.3.3 QTL克隆

QTL克隆的策略之一是首先将控制同一性状的多个基因分解开来，然后采用常规的或有待发展的专门的基因克隆方法将其一一克隆出来。QTL作图为将控制同一性状的多个QTL分解成为单个的可克隆基因提供了一种可能性。另一种方法是通过歧化选择(Disruptive selection)将多个数量基因区分开来。在基因克隆方法中，以图谱基础的基因克隆技术是克隆未知产物基因的有效方法。对于QTL的克隆，还有一些特别的方法可以考虑，其中包括：根据被用于杂交的某克隆顺序等位基因的直接效应进行克隆，利用转座子的插入突变进行克隆，以cDNA序列分析为基础的克隆，根据QTL与质量基因的同源性进行克隆等。为此，还要开展大量与QTL克隆有关的基础研究。有关内容详见第11章。

1.3.4 QTL序列分析

QTL精细结构和功能的研究有待于对影响同一性状的多个QTL进行详尽的序列分析。这类分析无疑将给我们对QTL的认识和了解带来革命性的变化。QTL序列分析也是我们认识多个QTL之间相互关系的基础。可以预期，未来基因组研究特别是分辨率最高的物理图谱——DNA序列的测定，将会最终导致数量性状遗传改良计划的深刻变革。

1.4 分子数量遗传学的发展前景与展望

分子数量遗传学以动植物的重要数量性状为研究对象，这就决定它既不是分子遗传学理论和方法在数量性状上的简单应用，也不是数理统计方法向分子遗传学的简单引入。在理论上，我们可将分子生物学领域发展的各种分子克隆技术应用于数量性状的遗传操纵。但在实践上，由于数量性状的评价对于群体的依赖以及外界环境对数量表型的干扰，因而需要发展与之相适应的研究方法。

为定向地对动植物进行遗传操纵和改良，育种家需要了解与重要数量性状表达有关的遗传因子数、遗传因子在染色体上的分布，各遗传因子对性状表达所起作用的相对大小。RFLP等分子标记的发展为上述问题的解决提供了现实可能性。分子标记系统的建立，使育种家可以根据亲本来源追踪染色体片段的传递，并从杂种后代中鉴定带有目标基因座位的个体。而且通过跟踪染色体片段而不是特定的基因，可能选择到带有尽可能少的非必需DNA的个体。这种借助于分子标记进行“全基因组选择”(Whole genome selection; Young和Tanksley 1989)的途径是基于以下两个重要概念：即育种决策是以染色体片段的长度和分布为依据的；目标基因附近的非理想片段(连锁累赘)以及在非连锁染色体上的非必需片段都可以通过对目标基因座位的选择加以去除。

分子标记尤其是RFLP的应用已被证明是研究基础和应用植物遗传学的有用工具。迄今RFLP已被用于揭示不同物种的基因组结构，近缘物种间的进化关系，多基因性状的遗传结构，QTL和孟德尔基因座位之间的相互关系等。目前，存在两个主要障碍阻止了RFLP等分子标记在这一领域的广泛应用：难以分析大量的样本以及难以确定QTL与单个“可克隆”基因之间的关系。现在RFLP分析的主要技术包括基因组DNA的纯化、DNA的酶切、DNA片段的电泳分离、将DNA从凝胶上转移到适当的膜上、同位素标记和分子杂交。而在理论上，只有其中两个步骤是必需的，即基因组DNA的分离和序列变异的测定。以DNA多聚酶链式反应 (PCR)为基础的分子标记技术基本上只包含了上述两个必要的步骤，从而使分析大大简化，可分析较为大量的样本。第二个问题是无法将 RFLP 分析所检测到的QTL与将来在遗传工程方法中可能克隆和操纵的单基因联系起来，近年来，新的更有效的基因克隆技术似乎可以解决 QTL 克隆的难题。遗传工程在植物改良上的推广，可望最终处理象产量、株高、生育期和香味等多基因遗传的重要经济性状。分子标记的应用已经和正在对我们有关多基因性状的理解产生重大影响，并将极大地简化采用传统育种方法对这些性状所实施的操纵。

只是一些年前，同工酶还是唯一的分子标记。现在我们已能生产和利用成千上万有丰富信息的各种分子标记。我们完全有理由相信，未来我们可能利用与之完全不同的全新分子标记。RCR 技术的发展使得不依赖于限制性酶或RFLP 但能检测到DNA 序列变异的方

法进入实用阶段。随着分子遗传学对于质量基因认识的进一步完善以及分子克隆技术的进一步发展，随着一些生物DNA全序列分析的逐步实施，分子数量遗传学也将进入一个全新的发展时期。

第 2 章　数量性状的生统遗传学基础

分子数量遗传学既涉及经典的数量遗传学理论和方法，又与分子遗传学及分子生物学密切相关。作为后续章节的基础，本章将着重介绍连续变异的遗传理论及其在育种中的应用。有一定数量遗传学基础的读者可直接进入后续章节。

2.1　数量性状及其遗传特点

2.1.1　数量性状与数量遗传学

遗传学上通常把生物体所表现的形态特征和生理特性等称为性状。根据性状在生物群体内个体间的差异，可以将其区分为质量性状和数量性状两大类。群体内的个体可以明确地分类和截然划分、表现为不连续性变异的性状称为质量性状 (Qualitative character)。例如，孟德尔所研究的豌豆子粒的形状（圆形与皱缩）、子叶的颜色（黄色与绿色）等。而群体内的个体不容易划分为明显的类别、表现为连续性变异的性状称为数量性状 (Quantitative character)，又称为计量性状 (Metric character)。例如，植物的种子重量、单株产量，人的身高、体重等。典型的质量性状和数量性状的区别特征如表 2.1。

表 2.1　质量性状和数量性状的区别特征

特　征	质量性状	数量性状
1. 所涉及的基因数目	一个或少数几个	几个到多个
2. 对环境的反应	不易受环境的影响	对环境变化极为敏感
3. 变异类型	种类上的变化	数量上的变化
4. 性状分布	不连续（间断）的多峰分布	连续的近似正态分布

数量遗传学是研究生物数量性状或连续变异的遗传规律的科学，它是遗传学的一个分支，在其发展历史上曾用过多种名称，如遗传数学(Mathematics of Heredity)，生统遗传学 (Biometrical Genetics)、遗传统计学 (Genetical Statistics)、数理遗传学 (MathematicalGenetics)、数量遗传学 (Quantitative Genetics)。现在一般倾向于采用数量遗传学这一名称。

2.1.2　群体及其遗传参数

在遗传学上把一群可以相互交配的个体所组成的集团称为群体 (Population)。这个集团在统计学上称为总体。在理论上一个群体由无限多个个体所组成。数量遗传学以生

物群体为研究对象，通常从群体中随机抽取一部分个体，以这些个体为样本 (Sample)，通过对样本的研究，获得一系列有关样本的统计数 (Statistic) 来估计群体的特征参数 (Parameter)，由此推测样本所来自的群体的特性。样本的随机性受到遗传的和非遗传的因素的影响。样本统计数和群体(总体)参数决定于群体的概率分布 (Distribution)。数量性状通常表现为连续变异的正态分布。

在数量遗传研究中，一般采用统计分析方法来估算遗传群体的统计参数，如平均数 (Mean)、方差 (Variance) 和协方差 (Covariance)。假定有一个品种，其基因型是纯合的。对于产量这个数量性状而言，每个单株的遗传组成是完全一样的。如果不存在影响每个单株产量的环境效应，则所有的单株将有相同的产量表现，其值等于群体平均数。但实际上，单株产量除了受基因型影响之外，还在很大程度上受环境因素(如温度、光照、水分、营养等)的影响，因而群体中每个单株的产量表现不尽相同，呈现连续的变异。在杂种后代分离群体中，每一个体可能具有不同的基因型，个体性状的表现是由基因分离和环境效应随机影响的综合作用造成的。因此对特定群体数量性状的遗传分析，必须采用统计分析的方法，将群体中个体间的变异分解为由遗传因素引起的变异和由环境因素引起的变异。

在统计学上，群体的变异程度常用方差来度量。表现为连续变异的数量性状，其性状值呈正态分布。反映正态分布特征的两个参数是群体平均数 μ 和方差 σ^2。常用 $N(\mu, \sigma^2)$ 来表示正态分布具有群体平均数 μ 和方差 σ^2。

在上述产量性状中，μ 度量了品种的基因型值，σ^2 度量环境效应所造成的变异程度。任何一个品种(群体)可以种植(或包含)的个体数在理论上是无穷的，我们不可能观测分析所有的个体，因而无法得知 μ 和 σ^2 的大小。采用统计分析的方法，只需从群体中随机抽取一些个体作为样本，通过样本植株的观测分析，便可以估算群体平均数 μ 和方差 σ^2。如果调查了 n 个个体，各个体的产量测定值为 x_1，x_2，…，x_i，…，x_n，群体平均数 μ 和方差 σ^2 可分别由样本均值 $\bar{x}$ 和样本方差 s^2 估计为

$$\hat{\mu} = \bar{x} = (x_1 + x_2 + \dots + x_i + \dots + x_n) / n = (1/n) \sum_{i=1}^{n} x_i$$

和

$$\hat{\sigma}^2 = s^2 = [(x_1 - \hat{\mu})^2 + (x_1 - \hat{\mu})^2 + \dots + (x_i - \hat{\mu})^2 + \dots + (x_n - \hat{\mu})^2]/(n-1)$$

$$= [\sum_{i=1}^{n} (x_i - \hat{\mu})^2]/(n-1)$$

样本平均数 $\bar{x}$ 和方差 s^2 是群体均值 μ 和方差 σ^2 的无偏估计。

由于存在基因的连锁和一因多效等，生物体的不同数量性状之间常存在不同程度的相互关联。在统计分析中常用协方差来度量这种相互关联的变异程度。如果某遗传群体有相互关联的数量性状 x 和 y，这两个性状的协方差 σ_{xy} 可用样本协方差 Cov_{xy} 估计为

$$\hat{\sigma}_{xy} = \mathrm{Cov}_{xy} = [\sum_{i=1}^{n} (x_i - \hat{\mu_x})(y_i - \hat{\mu_y})] / (n-1)$$

其中 x_i 和 y_i 的第 i 个观测值，$\hat{\mu_x}$ 和 $\hat{\mu_y}$ 则分别由两个性状的样本均值估计。

下面以 East(1915) 关于烟草花冠长度的研究为例说明数量性状的遗传特点。以两个烟草品种作亲本 P_1和 P_2，杂交产生杂种F_1，F_1自交产生F_2和F_3。根据群体中不同花冠长度的个体的出现频率，可以列出次数分布表(表2.2)。 反应各群体表型特征的统计数是对根据所观测的个体数(样本容量)计算出的每一群体的平均花冠长度(样本均值)及其变异程度(以样本方差表示)。 由F_2代的每一个体自交产生下一代F_3的株系，对每一F_3株系也可计算出花冠长度的均值和方差(表2.3)。上下代之间花冠长度的相关性程度(以相关系数 r

表 2.2 两个烟草品种及杂种后代花冠长度(mm) 的次数分布 (East 1915)

群	组 中 值 (mm)																							样本	样本	样本
体	34	37	40	43	46	49	52	55	58	61	64	67	70	73	76	79	82	85	88	91	94	97	100	容量 n	均值 $\bar{x}$	方差 s^2
P_1																			13	45	91	9	1	169	93.1	4.9
P_2	1	21	140	39																				211	40.4	2.3
F_1								4	10	41	75	40	3											173	63.5	7.9
F_2							3	9	18	47	55	93	75	60	43	25	7	8	1					444	68.8	41.6

表 2.3 烟草 9 个F_3家系花冠长度的表型均值和方差

株 系	1	2	3	4	5	6	7	8	9
F_2的花冠长度	46	50	50	60	72	77	80	81	82
F_3家系均值	53.5	50.2	53.0	56.3	73.1	73.0	74.0	76.3	80.2
F_3家系方差	13.9	10.0	9.2	16.5	14.5	24.9	23.4	25.5	22.6

$r_{F_3F_2}=0.9726^{**}$ $b_{F_3.F_2}=0.7776^{**}$

和回归系数 b 表示)，也可以反映该性状的主要遗传特征。

根据表2.2 和表2.3，可以总结出数量性状的几个遗传特点：

1. 无论分离群体(F_2、F_3) 还是非分离群体(P_1、P_2、F_1)，个体间数量性状表现型均呈连续变异；

2. 各群体的变异呈正态或近似正态分布；

3. 各群体以F_2 的变异最大，F_3 各家系次之；

4. F_3 家系均值与亲代 F_2 个体表型值之间存在极显著正相关。

数量性状连续变异的形成原因可以用多基因假说(Multigene hypothesis) 来解释。多基因假说认为数量性状是能够遗传的，它受多基因控制，这些基因存在于细胞核内的染色体上，它们在世代相传中，和控制质量性状的主基因一样，按孟德尔方式遗传，表现分离和重组，符合遗传的三大法则。但这些基因的效应较小，各基因的作用相似，等位基因间无明显的显隐性关系，位点间效应呈相加式，且易受外界环境条件的影响。后来数量遗传学的发展还证实了等位基因间的显隐性，不同基因座位间的连锁和互作等。

2.2 群体平均数及其遗传分量

2.2.1. 平均数的计算公式

在概率论 (Theory of probability) 中,均值表示为随机变量 x 的数学期望:

$$E(x) = \sum_{i=1}^{n} p_i x_i$$

其中 p_i 为 x_i 出现的概率。

例如: 表2.4 列出了50人的身高(cm)观测结果,包括分组、每组人数和占总人数的比例(概率),据此可以计算均值 $\bar{x}$ 如下:

$\bar{x}$ = (166+166+167+167+167+167+168+...+174+174)/50=8500/50=170 (cm)

E(x)= 0.04×166+0.08×167+0.12×168+...+0.04×174 = 170 (cm)

表 2.4 50 人的身高数据

身高分组 (x_i)	人数	概率 (p_i)
166	2	0.04
167	4	0.08
168	6	0.12
169	8	0.16
170	10	0.20
171	8	0.16
172	6	0.12
173	4	0.08
174	2	0.04
总计	50	1.00

2.2.2 群体表型值的分解

对群体中的个体进行数量性状的表型测量可以得到反映该性状表现程度的一个数量值,我们称其为表现型值(Phenotypic value, p),简称表型值。1909年, W. Johannsen通过对不同纯系的菜豆粒重变异的研究,发现性状的表现型变异可以归因于纯系间的变异和纯系内的变异。前者可以稳定遗传,属于遗传变异;后者是环境因素造成的,不能遗传,故称环境变异。如果某性状的变异受到遗传(基因型)效应(Genotypic effect, g)和环境中随机误差效应(又称为环境差值, Environmental deviation, e)的影响,则群体中某一个体 i 的表现型值(Phenotypic value, p) 可以用一个简单的线性公式来表示

$$p = \mu + g + e$$

即基因型效应 g 和环境效应 e 共同作用的结果使群体中某一个体的表型值偏离群体均值 μ。

对于一个群体而言，e 是均值为0，方差为 σ_e^2 的正态随机变量。一个个体的基因型效应加上群体均值即得到该个体的基因型值(Genotypic value，G)，它是表现型值中由遗传因素控制的部分。

对于具有 N 个个体的大群体，其群体的平均表现型值 P 即为群体均值：

$$P = [\sum_{i=1}^{N}(\mu + g_i + e_i)] / N = \mu + [\sum_{i=1}^{N} g_i + \sum_{i=1}^{N} e_i] / N = \mu$$

显然群体的平均基因型值也等于群体均值。

2.2.3 基因效应 (Gene effect)

假定基因座位A-a具有一对等位基因 A 和 a，其中 A 称为增效等位基因，a称为减效等位基因，其大小写分别表示其对基因型值的增加或降低作用。则对于在该基因座位上有差异的两个亲本 AA和 aa，其杂种后代群体中存在 AA、Aa 和 aa 三种基因型。我们将两种纯合体 AA 和 aa的基因型效应之差的一半定义为该座位的基因加性效应(Additive effect)，用 d 表示，即

$$d = (g_{AA} - g_{aa})/2$$

因此，AA 和 aa 基因型的效应分别为 $+d$ 和 $-d$。

对于影响同一性状的 k 个基因座位，其总的基因加性效应为

$$\sum_{i=1}^{k} d_i = (d_1 + d_2 + \ldots + d_k) = [d]$$

杂合体Aa的基因型效应偏离两种纯合体的平均基因型效应的程度定义为基因的显性效应 (Dominance effect)，用 h 表示：

$$h = g_{Aa} - (g_{AA} + g_{aa})/2$$

其中 $(g_{AA} + g_{aa})/2$ 为两种纯合体(或亲本)的平均基因型效应。

对于影响同一性状的 k 个基因座位，其总的基因显性效应为

$$\sum_{i=1}^{k} h_i = (h_1 + h_2 + \ldots + h_k) = [h]$$

上述加性效应和显性效应的估计式中的基因型效应也可以用相应的基因型值来代替。如考虑同一基因型所组成的群体，也可用其相应的表型均值代替。此时

$$d = (P_{AA} - P_{aa})/2$$

$$h = P_{Aa} - (P_{AA} + P_{aa})/2$$

$$= P_{Aa} - m$$

其中 m 为两种纯合体(亲本)的平均表型值，也称中亲值。

单基因座位和多基因座位下，各类基因型的基因型效应和基因型值列于表2.5

根据 h 的大小和方向的不同，Aa 的基因型效应和基因型值在 AA 和aa 的数轴上有7 种可能的位置(图2.1 和表2.6)。

表 2.5　单基因座位和多基因座位情形下的基因型、基因型效应和基因型值

单基因座位			多基因座位		
基因型	基因型效应	基因型值	基因型	基因型效应	基因型值
AA	d	$m+d$	$A_1A_1A_2A_2\cdots A_kA_k$	$[d]$	$m+[d]$
Aa	h	$m+h$	$A_1a_1A_2a_2\cdots A_ka_k$	$[h]$	$m+[h]$
aa	$-d$	$m-d$	$a_1a_1a_2a_2\cdots a_ka_k$	$-[d]$	$m-[d]$

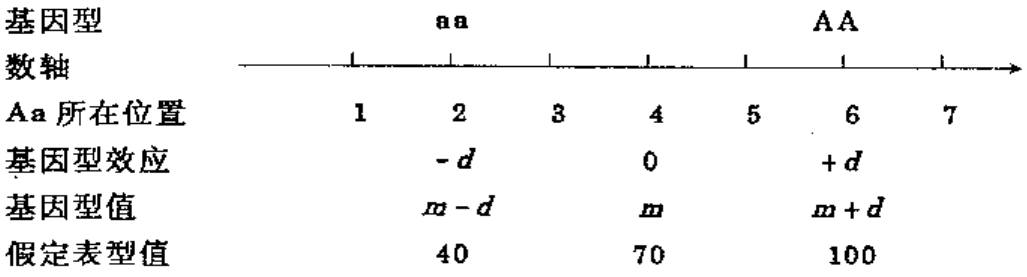


图 2.1　基因显性效应的大小和方向对杂合体的影响

表 2.6　基因显性效应的大小和方向对杂合体表现型的影响

Aa 所在位置	表型值	显性方向和大小	
1	<40	负向超显性	$\lvert h/d \rvert > 1$
2	40	负向完全显性	$\lvert h/d \rvert = 1$
3	40<Aa<70	负向部分显性	$0 < \lvert h/d \rvert < 1$
4	70	无显性	$h=0$
5	70<Aa<100	正向部分显性	$0<h/d<1$
6	100	正向完全显性	$h/d=1$
7	>100	正向超显性	$h/d>1$

2.2.4　近交群体世代平均数的遗传分量

经近亲交配所产生的群体称为近交群体。常见的近交群体包括自交群体、回交群体、同胞交配群体等。

由两个纯系亲本 P_1 和 P_2 杂交产生 F_1。这三个群体为基因型一致的非分离群体，其群体基因型值（由群体均值估计）分别为：

群体均值	单基因座位	多基因座位
P_1	$m + d$	$m + [d]$
P_2	$m - d$	$m - [d]$
F_1	$m + h$	$m + [h]$

由该F_1自交所产生的F_2，其均值（单基因座位）可计算如下：

基因型种类	频率（p_i）	基因型值（x_i）	$p_i x_i$
AA	1/4	$m + d$	$(m+d)/4$
Aa	1/2	$m + h$	$(m+h)/2$
aa	1/4	$m - d$	$(m-d)/4$
总计（Σ）	1		$m + h/2$

$$\overline{F}_2 = \sum_{i=1}^{3} p_i x_i = 1/4\ (m+d) + 1/2(m+h) + 1/4(m-d) = m + h/2$$

对于k对独立基因

$$\overline{F}_2 = m + [h]/2$$

在自交群体中，两类纯合体的基因型效应相互抵消，只有杂合体对群体均值的大小有影响。在自交第t代，杂合体在群体中所占的比例为$(1/2)^t$，其均值为：

$$\overline{F}_t = m + (1/2)^{t-1}\ [h]$$

F_1或杂种后代与亲本之一再杂交称为回交，由此所产生的群体称为回交群体（backcross population），记F_1与P_1回交为B_1，F_1与P_2回交为B_2，则有

$$\overline{B}_1 = m + 1/2\ [d] + 1/2\ [h]$$

$$\overline{B}_2 = m - 1/2\ [d] + 1/2\ [h]$$

向大值亲本连续回交t次记为B_{1t}，向小值亲本连续回交t次记为B_{2t}，回交群体中的杂合体频率为$(1/2)^t$，纯合体频率为$1 - (1/2)^t$，因此

$$\overline{B}_{1t} = m + [1 - (1/2)^t]\ [d] + (1/2)^t\ [h]$$

$$\overline{B}_{2t} = m - [1 - (1/2)^t]\ [d] + (1/2)^t\ [h]$$

类似地可以获得其它近交群体均值的遗传组成分量（表2.7）

2.3 群体方差及其遗传分量

2.3.1 方差的概念及其计算

在概率论中，随机变量x的方差定义为

$$\sigma_x^2 = \Sigma(x^2) - (\Sigma x)^2 = \Sigma p_i x_i^2 - (\Sigma p_i x_i)^2$$

方差所反映的是群体内个体间的变异程度。表2.8给出了两个人群x和y各50个个体的身高数据。据此可以计算出身高的样本方差Var x和Var y。

表 2.7　近交群体平均数的遗传组成分量（加性-显性模型）

群　体	一对等位基因			k 对等位基因		
	m	d	h	m	$[d]$	$[h]$
P_1	1	1	0	1	-1	0
P_2	1	-1	0	1	-1	0
F_1	1	0	1	1	0	1
F_2	1	0	1/2	1	0	1/2
F_s	1	0	1/4	1	0	1/4
F_t	1	0	$(1/2)^{t-1}$	1	0	$(1/2)^{t-1}$
B_1	1	1/2	1/2	1	1/2	1/2
B_2	1	-1/2	1/2	1	-1/2	1/2
B_{1t}	1	$1-(1/2)^t$	$(1/2)^t$	1	$1-(1/2)^t$	$(1/2)^t$
B_{2t}	1	$-[1-(1/2)^t]$	$(1/2)^t$	1	$-[1-(1/2)^t]$	$(1/2)^t$
B_1s	1	1/2	1/4	1	1/2	1/4
B_2s	1	-1/2	1/4	1	-1/2	1/4
$F_2\times P_1$	1	1/2	1/2	1	1/2	1/2
$F_2\times P_2$	1	-1/2	1/2	1	-1/2	1/2
$F_2\times F_1$	1	0	1/2	1	0	1/2
F_2bip	1	0	1/2	1	0	1/2
B_1bip	1	1/2	3/8	1	1/2	3/8
B_2bip	1	-1/2	3/8	1	-1/2	3/8

s表示自交；bip表示成对交

表 2.8　两个人群身高的数据及其方差的计算

身高分组	样本　x				样本　y			
(x,y)	人数	p_i	p_ix_i	$p_ix_i^2$	人数	p_i	p_iy_i	$p_iy_i^2$
166	2	0.04	0.04×166	0.04×166^2				
167	4	0.08	0.08×167	0.08×167^2				
168	6	0.12	0.12×168	0.12×168^2				
169	8	0.16	0.16×169	0.16×169^2	10	0.20	0.20×169	0.20×169^2
170	10	0.20	0.20×170	0.20×170^2	30	0.60	0.60×170	0.60×170^2
171	8	0.16	0.16×171	0.16×171^2	10	0.20	0.20×171	0.20×171^2
172	6	0.12	0.12×172	0.12×172^2				
173	4	0.08	0.08×173	0.08×173^2				
174	2	0.04	0.04×174	0.04×174^2				
Σ	50	1.00	170.0	28904	50	1.00	170.0	28900.4

$$\text{Var}\,x = \sum p_i x_i^2 - (\sum p_i x_i)^2 = 28904 - 170^2 = 4 \qquad \text{Var}\,y = \sum p_i y_i^2 - (\sum p_i y_i)^2 = 28900.4 - 170^2 = 0.4$$

$$\bar{x} = 170.0 \qquad \bar{y} = 170.0$$

2.3.2 群体性质和方差分量

采用类似于表现型值的分解方法可以将表型方差(Phenotypic variance, σ_P^2)分解为由基因型作用所引起的基因型方差(Genotypic variance, σ_G^2)和由环境作用所引起的环境方差 (Environmental variance, σ_e^2), 即

$$\sigma_P^2 = \sigma_G^2 + \sigma_e^2$$

基因型方差 (σ_G^2) 可进一部分解为由基因的加性和显性作用所引起的方差(σ_A^2和σ_D^2)

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2$$

若群体内的个体具有相同的基因型即同质群体，各个体的基因型是一致的，其表型变异由环境因素引起，此时有

$$\sigma_P^2 = \sigma_e^2$$

2.3.3 近交群体的方差分量

纯系亲本和 F_1 为基因型一致的群体，故有

$$\sigma_{P_1}^2 = \sigma_{P_2}^2 = \sigma_{F_1}^2 = \sigma_e^2 = E$$

F_2 的基因型方差 $\sigma_{GF_2}^2$ (先考虑一对基因) 可根据表2.9 计算。

表 2.9 F_2群体遗传方差的计算

F_2基因型种类	频率(p_i)	基因型值	基因型效应(x_i)	$p_i x_i$	$p_i x_i^2$
AA	1/4	$m+d$	$+d$	$1/4\ d$	$1/4\ d^2$
Aa	1/2	$m+h$	h	$1/2\ h$	$1/2\ h^2$
aa	1/4	$m-d$	$-d$	$1/4\ (-d)$	$1/4\ (-d)^2$
Σ	1			$1/2h$	$1/2d^2 + 1/2h^2$

$$\sigma_{GF_2}^2 = \sum p_i x_i^2 - (\sum p_i x_i)^2 = 1/2\,d^2 + 1/2\,h^2 - (1/2\,h)^2 = 1/2\ d^2 + 1/4\ h^2$$

当性状由 k 对独立基因控制时:

$$\sigma_{GF_2}^2 = 1/2\ \sum d_i^2 + 1/4\ \sum h_i^2 = 1/2\,D + 1/4\,H$$

其中

$$D = \sum_{i=1}^{k} d_i^2 \qquad H = \sum_{i=1}^{k} h_i^2$$

F_2 的表型方差 $\sigma_{PF_2}^2$为:

$$\sigma_{PF_2}^2 = \sigma_{GF_2}^2 + \sigma_{eF_2}^2 = 1/2\ D + 1/4\ H + E_{F_2}$$

F_2群体遗传方差（先考虑一对基因）可根据表2.10计算。

表2.10　F_2群体遗传方差的计算

F_2 基因型	频率(p_i)	基因型效应（x_i）	$p_i x_i$	$p_i x_i^2$
AA	3/8	$+d$	$3/8\ d$	$3/8\ d^2$
Aa	1/4	h	$1/4\ h$	$1/4\ h^2$
aa	3/8	$-d$	$3/8(-d)$	$3/8\ (-d)^2$
Σ	1		$1/4\ h$	$3/4\ d^2 + 1/4\ h^2$

$$\sigma^2_{gF_2} = 3/4\,d^2 + 1/4\,h^2 - (1/4\,h)^2 = 3/4\,d^2 + 3/16\,h^2 \rightarrow 3/4\,D + 3/16\,H$$

F_2的遗传方差 $\sigma^2_{gF_2}$可分解为由 F_1配子的分离所引起的遗传方差 $\sigma^2_{gF_{2(1)}}$（一阶方差，即 F_2家系平均数的遗传方差）和由 F_2配子的分离所引起的遗传方差 $\sigma^2_{gF_{2(2)}}$（二阶方差，F_2家系内方差的平均）。这两类方差可按表2.11进行分解。

表 2.11　F_2群体两类遗传方差的分解

F_2		F_3		$p_i x_i$	$p_i x_i^2$
基因型	频率(p_i)	家系平均数(x_i)	家系内方差(y_i)		
AA	1/4	d	0	$1/4\,d$	$1/4\,d^2$
Aa	1/2	$1/2\,h$	$1/2\,d^2 + 1/4\,h^2$	$1/2(1/2\,h)$	$1/2(1/2\,h)^2$
aa	1/4	$-d$	0	$1/4(-d)$	$1/4(-d)^2$
Σ	1		$1/2\,d^2 + 1/4\,h^2$	$1/4\,h$	$1/2\,d^2 + 1/8\,h^2$

$$\sigma^2_{gF_{2(1)}} = \Sigma p_i x_i^2 - (\Sigma p_i x_i)^2$$
$$= 1/2\,d^2 + 1/8\,h^2 - (1/4\,h)^2$$
$$= 1/2\,d^2 + 1/16\,h^2$$
$$= 1/2\,D + 1/16\ H \quad (\text{对于 } k \text{ 对独立基因})$$

$$\sigma^2_{gF_{2(2)}} = \Sigma p_i y_i = 1/4\ (0) + 1/2\ (1/2\,d^2 + 1/4\,h^2) + 1/4\ (0)$$
$$= 1/4\,d^2 + 1/8\,h^2$$
$$= 1/4\,D + 1/8\,H \quad (\text{对于 k 对独立基因})$$

而　$\sigma^2_{gF_2} = \sigma^2_{gF_{2(1)}} + \sigma^2_{gF_{2(2)}}$

对于自交第 t 代，其遗传方差为：

$$\sigma^2_{gF_t} = \left[\ 1 - (1/2)^{t-1}\ \right] D + \frac{2^{t-1} - 1}{4^{t-1}}\ H$$

由此可得 F_t 代比 F_{t-1} 代的加性方差的增量为：

$$\Delta D = \frac{2^{t-1}-1}{2^{t-1}} D - \frac{2^{t-2}-1}{2^{t-2}} D = (1/2)^{t-1} D$$

F_t 代比 F_{t-1} 代的显性方差的减少量为:

$$\Delta H = \frac{2^{t-2}-1}{4^{t-2}} H - \frac{2^{t-1}-1}{4^{t-1}} H = \frac{2^{t-1}-3}{4^{t-1}} H$$

B_1 群体的遗传方差 σ^2_{gB1} (先考虑一对基因)可根据表2.12计算。

表 2.12 回交群体B_1遗传方差的计算

基因型种类	频率(p_i)	基因型效应(x_i)	$p_i x_i$	$p_i x_i^2$
AA	1/2	d	$1/2\,d$	$1/2\ \ d^2$
Aa	1/2	h	$1/2\,h$	$1/2\ \ h^2$
Σ	1		$1/2\ \ (d+h)$	$1/2\,d^2 + 1/2\,h^2$

$$\sigma^2_{gB1} = 1/2\,d^2 + 1/2\,h^2 - [1/2\ \ (d+h)]^2 = 1/4\,d^2 - 1/2\,dh + 1/4\,h^2$$
$$= 1/4\,D - 1/2\,F + 1/4\,H \quad \text{(对于 } k \text{ 对独立基因)}$$

其中 $F = \sum_{i=1}^{k} d_i h_i$

同理可得:

$$\sigma^2_{gB2} = 1/4\,D + 1/2\,F + 1/4\,H$$

因而有 $\sigma^2_{gB1} + \sigma^2_{gB2} = 1/2\,D + 1/2\,H$

一些常见近交群体的遗传方差分量列于表 2.13。

2.3.4 群体性质和协方差分量 (Covariance components)

概率论中两个随机变量x和y的协方差定义为:

$$\mathrm{Cov}_g(x,y) = \Sigma\ p_i x_i y_i - (\Sigma\, p_i x_i)(\Sigma\, p_i y_i)$$

对于同质群体 P_1、P_2等,由于上下代不存在由于遗传因素引起的相关性,上下代之间的表型协方差 Cov_p 就等于环境协方差 Cov_e:

$$\mathrm{Cov}_p(P_1,P_1) = \mathrm{Cov}_e(P_1,P1)$$
$$\mathrm{Cov}_p(P_2,P_2) = \mathrm{Cov}_e(P_2,P_2)$$

对于异质群体,上下代之间的表型协方差可分解为由于遗传原因引起的遗传协方差及由于环境因素引起的环境协方差:

$$\mathrm{Cov}_p = \mathrm{Cov}_g + \mathrm{Cov}_e$$

F_2 和 F_3 之间的遗传协方差$\mathrm{Cov}_g(F_2,F_3)$(先考虑一对基因) 可按表2.14计算。

$$\mathrm{Cov}_g(F_2,F_3) = \Sigma\ p_i x_i y_i - (\Sigma\, p_i x_i)(\Sigma\, p_i y_i)$$
$$= 1/2\,d^2 + 1/4\,h^2 - (1/2\,h)\ \ (1/4\,h)$$

表 2.13　常见群体的遗传方差分量（加性-显性模型）

群　体	符　号	D	H	F
F_2	σ^2_{gF2}	1/2	1/4	
F_3	σ^2_{gF3}	3/4	3/16	
F_4	σ^2_{gF4}	7/8	7/64	
F_t	σ^2_{gFt}	$1-(1/2)^t$	$(2^{t-1}-1)/4^{t-1}$	
B_1	σ^2_{gB1}	1/4	1/4	-1/2
B_2	σ^2_{gB2}	1/4	1/4	1/2
	$\sigma^2_{gB1}+\sigma^2_{gB2}$	1/2	1/2	
B_1S	σ^2_{gB1S}	1/4	1/8	
B_2S	σ^2_{gB2S}	1/4	1/8	
S_2	σ^2_{gS2}	1/2	1/4	
S_3	σ^2_{gS3}	1/2	1/4	
$F_2\times P_1$		1/8	1/8	-1/4
$F_2\times P_2$		1/8	1/8	1/4
$F_2\times F_1$		3/8	1/4	
F_2bip		1/4	3/16	
B_1bip		1/4	3/16	-1/4
B_2bip		1/4	3/16	1/4

S_2、S_3 为同胞交配群体

表2.14　F_2和F_3之间遗传协方差的计算

	F_2		F_3家系均值			
基因型	频率(p_i)	效应(x_i)	(y_i)	p_ix_i	p_iy_i	$p_ix_iy_i$
AA	1/4	$+d$	d	$1/4\,d$	$1/4\,d$	$1/4\,d\times d$
Aa	1/2	h	$1/2\,h$	$1/2\,h$	$1/2\times1/2\,h$	$(1/2\,h)\times(1/2)\,h$
aa	1/4	$-d$	$-d$	$1/4(-d)$	$1/4(-d)$	$1/4\ (-d)(-d)$
Σ	1			$1/2\,h$	$1/4\,h$	$1/2\,d^2+1/4\,h^2$

$$= 1/2\,d^2 + 1/8\,h^2$$

$$= 1/2\,D + 1/8\,H \qquad \text{(对于 } k \text{ 对独立基因)}$$

同理可得：

$$\mathrm{Cov}_g(F_1,F_2) = 1/4\,d\times h + 1/2\,h\times h + 1/4(-d)\times h - (h)\ (1/2\,h) = 0$$

$$\mathrm{Cov}_g(F_3,F_4) = 3/4\,D + 3/32\,H$$

$$\mathrm{Cov}_g(F_t, F_{t+1}) = \frac{2^{t-1}-1}{2^{t-1}}D + \frac{2^{t-1}-1}{2^{2t-1}}H$$

2.4 数量遗传模型及其检验

可以将模型 (Model) 简单地定义为科学规律的数学表达式。例如人的身高 H 与体重 W 之间在一定范围内存在下列数学模型

$$W\text{（公斤）} = H\text{（厘米）} - 100$$

数量遗传模型就是数量性状的遗传表现与各种遗传效应之间的函数关系。

2.4.1 加性-显性遗传模型 (Additive-dominance model)

加性-显性遗传模型建立的基础是世代群体均值之间的相互关系。在本节以前所介绍的群体均值、方差和协方差，都只考虑了基因的加性和显性效应，而未考虑基因座位间的互作和基因连锁，即它们是在加性-显性模型假定下获得的。

由世代均值

$$P_1 = m + [\ d\ ]$$
$$P_2 = m - [\ d\ ]$$
$$F_1 = m + [\ h\ ]$$
$$F_2 = m + 1/2\ [\ h\ ]$$

可得

$$F_2 - 1/2\ (P_1 + P_2) = 1/2\ [\ F_1 - 1/2\ (P_1 + P_2)\ ]$$

即 $m + 1/2\ h - 1/2\ (m + d + m - d) = 1/2\ [\ m + h - 1/2\ (m + d + m - d)\ ]$

所以

$$F_2 = 1/4\ P_1 + 1/2\ F_1 + 1/4\ P_2$$

类似地有:

$$F_3 = 1/4\ P_1 + 1/4\ P_2 + 1/2\ F_2 = 3/8\ P_1 + 3/8\ P_2 + 1/4\ F_1$$
$$B_1 = 1/2\ P_1 + 1/2\ F_1$$
$$B_2 = 1/2\ P_2 + 1/2\ F_1$$
$$\cdots\quad\cdots$$

在一个数量性状的遗传试验中，根据世代均值间的相互关系，我们可以检验所采用的遗传模型对这一性状是否适合。其原理就是比较供试群体的均值，检查在误差范围内上述世代均值之间的期望关系是否成立。其中最简单的检验方法是 ABC 尺度测验 (ABC Scaling test)。该检验的基础是在加性-显性模型下，以下三个等式的期望值为0:

$$A = 2\ B_1 - P_1 - F_1$$
$$B = 2\ B_2 - P_2 - F_1$$
$$C = 4\ F_2 - 2\ F_1 - P_1 - P_2$$

可以采用 t-测验来检验由世代平均值所得到的观测值 A、B、C与理论值 0 的差异显著性。例如，对于 A 尺度:

$$t_A = \frac{|A|}{(\mathrm{Var}_A)^{1/2}}$$

其中

$$\begin{aligned}\mathrm{Var}_A &= \mathrm{Var}(2B_1 - P_1 - F_1)\\ &= \mathrm{Var}(2B_1) + \mathrm{Var}(-P_1) + \mathrm{Var}(-F_1)\\ &= 4\ \mathrm{Var}(B_1) + \mathrm{Var}(P_1) + \mathrm{Var}(F_1)\end{aligned}$$

2.4.2 加性-显性-上位性模型

基因的非独立作用包括非等位基因互作和基因连锁。基因互作又称为上位性(Epistasis)。基因互作可分为两个基因座位所引起的二基互作及三个或更多基因座位所引起的多基因互作。

涉及 A_1-a_1 和 A_2-a_2 两个基因座位的二基因互作可分为以下几类:

两个基因座位加性效应之间的互作称为加性×加性互作，记为 $d_1 \times d_2$ 或 i_{12}; 一个基因座位的加性效应与另一个基因座位显性效应的互作称为加性×显性互作，记为 $d_1 \times h_2$、$d_2 \times h_1$ 或 j_{12}、j_{21};两个座位的显性效应的互作称为显性×显性的互作，记为 $h_1 \times h_2$ 或 l_{12}。

涉及两个基因座位时，有差异的一对纯合亲本间的 F_2 有 9 种可能的基因型，各基因型的效应列于表 2.15。其中各互作项的系数为两个主效项的系数的乘积。

表 2.15 两对基因互作条件下 F_2 群体 9 种基因型的效应

基因型	A_2A_2 (1/4)	A_2a_2 (1/2)	a_2a_2 (1/4)
A_1A_1 (1/4)	1/16 $(d_1+d_2+i_{12})$	1/8 $(d_1+h_1+j_{12})$	1/16 $(d_1-d_2-i_{12})$
A_1a_1 (1/2)	1/8 $(h_1+d_2+j_{21})$	1/4 $(h_1+h_2+l_{12})$	1/8 $(h_1-d_2-j_{21})$
a_1a_1 (1/4)	1/16 $(-d_1+d_2-i_{12})$	1/8 $(-d_1+h_2-j_{12})$	1/16 $(-d_1-d_2+i_{12})$

如果考虑三对基因互作，有关基因型的效应可表示如下:

$$g(A_1A_1A_2A_2A_3A_3) = d_1 + d_2 + d_3 + i_{12} + i_{13} + i_{23} + i_{123}$$

$$g(A_1A_1A_2A_2A_3a_3) = d_1 + d_2 + h_3 + i_{12} + j_{13} + j_{23} + j_{12/3}$$

$$g(A_1A_1A_2a_2A_3a_3) = d_1 + h_2 + h_3 + j_{12} + j_{13} + l_{23} + j_{1/23}$$

其中 i_{123} 为三个基因座位加性效应之间的互作效应，$j_{12/3}$ 为 1、2 两个座位的加性效应与第3座位的显性效应之间的互作效应，$j_{1/23}$ 为第1座位的加性效应与第2、3座位的显性效应之间的互作效应。

上述基因互作与经典的互作之间可找出相对应的关系 (Mather和Jinks 1982)。

在加性-显性-二基因互作模型下，几类常见群体的均值如下:

$P_1(A_1A_1A_2A_2) = m + d_1 + d_2 + i_{12}$

F_1(多对基因) $= m + [d] + [i]$

$P_2(a_1a_1a_2a_2) = m - d_1 - d_2 + i_{12}$

P_2(多对基因) = $m - [d] + [i]$

$F_1 = m + [h] + [l]$

$F_2 = m + 1/2\ [h] + 1/4\ [l]$

$B_1 = m + 1/2\ [d] + 1/2\ [h] + 1/4\ [i] + 1/4\ [j] + 1/4\ [l]$

$B_2 = m - 1/2\ [d] + 1/2\ [h] + 1/4\ [i] - 1/4\ [j] + 1/4\ [l]$

由此得到与平均数有关的 6 个参数 m、$[d]$、$[h]$、$[i]$、$[j]$ 和 $[l]$ 的估计式:

$m = 1/2\ P_1 + 1/2\ P_2 + 4\ F_2 - 2\ B_1 - 2\ B_2$

$[d] = 1/2\ P_1 - 1/2\ P_2$

$[h] = 6\ B_1 + 6\ B_2 - 8\ F_2 - F_1 - 3/2\ P_1 - 3/2\ P_2$

$[i] = 2\ B_1 + 2\ B_2 - 4F_2$

$[j] = 2\ B_1 - P_1 + P_2 - 2\ B_2$

$[l] = P_1 + P_2 + 2\ F_1 + 2\ F_2 - 4\ B_1 - 4\ B_2$

也可采用 t-测验检验各参数估计值与 0 的差异显著性，例如:

$t_{[d]} = |[d]| / (\mathrm{Var}_{[d]})^{1/2}$

$(\mathrm{Var}_{[d]})^{1/2} = (1/4\ \mathrm{Var}P_1 + 1/4\ \mathrm{Var}P_2)^{1/2}$

余类推。[i]、[j]和[l]三者之一显著即表明存在显著的上位性效应(非等位基因互作)，但不能说明是否符合加性-显性-上位性模型。

加性-显性-二基因互作模型不适合有三种可能原因: (1) 存在三基因或更多基因的互作; (2) 互作基因之间存在连锁; (3) 两者兼而有之。

可采用以下三式对多基因互作进行检验，当无互作时，X、Y、Z 的期望值为0:

$X = 1/2\ (P_1 - P_2) - (B_{11} + B_{12}) + (B_{21} + B_{22})$

$Y = F_1 - 1/2\ (P_1 + P_2) + (B_{11} - B_{12}) - (B_{21} - B_{22})$

$Z = (B_1 - B_2) - 2(\overline{B_1F_1} - \overline{B_2F_1})$

其中，B_{11}为$B_1 \times P_1$，B_{12}为$B_1 \times P_2$，B_{21}为$B_2 \times P_1$，B_{22}为$B_2 \times P_2$。

2.5 遗传变异的传递与选择

影响数量性状遗传变异传递和选择的因素包括: 控制性状的基因数目或有效因子数; 等位基因间显性效应的大小; 性状总变异中遗传方差(或加性遗传方差) 所占的比重等。

2.5.1 有效因子数(Number of effective factors)或最小基因数

对于质量性状，可根据分离群体中亲型个体出现的频率来推测与性状有关的基因数目。对于数量性状，经典的基因数目分析方法有四类:(1) 染色体分析法(Chromosome assay); (2) 理论与观测值比较法; (3) 统计参数法; (4) 基因型分析法 (Genotypic assay)。

Castle (1921) 根据F_2代的基因型频率按基因数目呈二项式分布的原理，提出了数量性状的基因数目估计方法。若存在 k 对等位基因的差异，则 F_2 代应有$2k+1$ 种基因型，而基因型的频率呈 $(1/2 + 1/2)^{2k}$ 的二项式分布。F_2群体的期望基因加性方差为二项分布的期望方差

$$npq = 2k \times 1/2 \times 1/2 = k/2 = \sigma^2_{dF_2}$$

假定同效等位基因在亲本间集中分布，每个基因座位的效应相等且为一个单位，则

$$\bar{P}_1-\bar{P}_2 = 2[d] = 2kd = 2k$$

而 $$(\bar{P}_1-\bar{P}_2)^2 = (2k)^2$$

F_2 群体基因加性方差与双亲差值的平方的理论比值和实际比值，在假定符合的情形下应相等，即

$$\frac{\sigma^2_{dF_2}}{(\bar{P}_1-\bar{P}_2)^2} = \frac{k/2}{(2k)^2}$$

解之得：

$$k = \frac{(\bar{P}_1-\bar{P}_2)^2}{8\sigma^2_{dF_2}}$$

其中 $\sigma^2_{dF_2}$ 一般由 $Var_{F_2}- Var_{eF_2}$来估计。上述估计式只有在纯加性遗传模型下，k 的估值才是无偏的。且在大多数情形下，假定不符合通常使 k 值被低估。

2.5.2 显性度 (Degree of dominance)

1. 由世代均值确定显性程度

一对基因时，显性度为 h/d；

k 对基因时，显性度为 $\Sigma h_i/\Sigma d_i = [h]/[d]$，此时又称为势能比(Potential ratio)。

2. 由世代方差确定显性程度

若存在 k 对等位基因，每个座位的基因效应相等，即

$$d_1 = d_2 = \dots = d_k = d;\ h_1 = h_2 = \dots = h_k = h$$

则 $$(H/D)^{1/2} = (\Sigma h_i^2/\Sigma d_i^2)^{1/2} = (kh^2/kd^2)^{1/2} = h/d$$

当 $d_i \neq d$，$h_i \neq h$ 时，$(H/D)^{1/2}$ 为平均显性度 (Degree of average dominance)。当$(H/D)^{1/2} = 0$时为无显性；当 $0 < (H/D)^{1/2} < 1$时为部分显性；当 $(H/D)^{1/2} > 1$ 时为超显性。

2.5.3 遗传率 (Heritability) 及其估计

遗传率被定义为 “观测方差中由遗传变异所引起的部分”。

可将遗传率进一步分为广义遗传率 (Heritability in broad sense)和狭义遗传率 (Heritability in narrow sense)。广义遗传率即“总遗传方差占表型方差的比率”：

$$h_B^2 = \frac{\sigma^2_G}{\sigma^2_P} = \frac{\sigma^2_G}{\sigma^2_G + \sigma^2_e}$$

而狭义遗传率为 “加性方差占表型方差的比例”：

$$h_N^2 = \frac{\sigma^2_A}{\sigma^2_P} = \frac{\sigma^2_A}{\sigma^2_A + \sigma^2_D + \sigma^2_e}$$

遗传率是指变异的传递率。假定人类的平均身高为 1.65 米，某对夫妇的身高为1.73，当遗传率 $h^2 = 0.55$ 即 55%时，预期该对夫妇的子代平均身高为 1.65 + (1.73-1.65) ×55% = 1.65 + 0.044 = 1.694 米。

1. 利用非分离群体及 F_2 估计 h_B^2

环境方差可由非分离世代估计:

$$\sigma^2_e = \sigma^2_{P_1} \text{ 或 } 1/2\ (\sigma^2_{P_1} + \sigma^2_{P_2}) \text{ 或 } 1/3\ (\sigma^2_{P_1} + \sigma^2_{P_2} + \sigma^2_{F_1})$$

$$\text{或 } 1/4\ (\sigma^2_{P_1} + 2\sigma^2_{F_1} + \sigma^2_{P_2})$$

$$= (\sigma^2_{P_1} \times \sigma^2_{P_2})^{1/2} \text{ 或 } (\sigma^2_{P_1} \times \sigma^2_{P_2} \times \sigma^2_{F_1})^{1/3} \text{ 等}$$

$$h_B^2 = \frac{\sigma^2_{F_2} - \sigma^2_e}{\sigma^2_{F_2}} = \frac{1/2\ D + 1/4\ H}{1/2\ D + 1/4\ H + E}$$

2. 利用上下代之间的回归和相关估算 h_b^2 和 h_r^2

F_3家系均值对F_2个体的回归 (Regression) 为

$$b_{\overline{F_3}, F_2} = \frac{\mathrm{Cov}_x\ (F_2,\ \overline{F_3})}{\sigma^2_{F_2}} = \frac{1/2\ D + 1/8\ H}{1/2\ D + 1/4\ H + E} = h_b^2$$

F_4家系群均值对F_3家系均值的回归为

$$b_{F_4, \overline{F_3}} = \frac{\mathrm{Cov}\ (\overline{F_3},\ F_4)}{\sigma^2_{\overline{F_{3(1)}}}} = \frac{1/2\ D + 1/32\ H}{1/2\ D + 1/16\ H + E} = h_b^2$$

为克服 $b > 1$ 时出现的难以解释的遗传率，Frey和Horner 1957) 提出标准单位遗传率——亲子代的原始数据除以自身的标准差后再计算回归系数，即原始数据的相关系数 r (Correlation coefficiency)。

F_3家系均值与 F_2 个体之间相关为

$$r_{\overline{F_3}, F_2} = \frac{\mathrm{Cov}\ (\overline{F_3},\ F_2)}{(\sigma^2_{F_2} \times \sigma^2_{\overline{F_3}})^{1/2}} = h_r^2$$

3. 利用回交世代估计 h_N^2

Warner (1952) 提出了应用 B_1、B_2和 F_2 群体估计 h_N^2的变异分解法。即由下述三个等式获得各遗传方差分量。

$$\sigma^2_{P_2} = 1/2\ D + 1/4\ H + E \tag{1}$$

$$\sigma^2_{B_1} = 1/4\ D + 1/4\ H - 1/2\ F + E \tag{2}$$

$$\sigma^2_{B_2} = 1/4\ D + 1/4\ H + 1/2\ F + E \tag{3}$$

(2) + (3) 得

$$\sigma^2_{B_1} + \sigma^2_{B_2} = 1/2\ D + 1/2\ H + 2E \tag{4}$$

由 (1) 和 (4) 得

$$D = 2\ [\ 2\sigma^2_{F_2} - (\sigma^2_{B_1} + \sigma^2_{B_2})\ ] = 1/2\ \sigma^2_A$$

$$h_N^2 = \frac{2\ \sigma^2_{F_2} - (\sigma^2_{B_1} + V_{B_2})}{\sigma^2_{F_2}} = \frac{1/2\ D}{1/2\ D + 1/4\ H + E}$$

2.5.4 直接选择 (Direct selection) 与选择响应 (Response to selection)

变异是物种形成和新品种选择的基础。选择造成群体中不同基因型的差异繁殖，导致群体基因频率的变化，从而导致所选择性状群体均值和方差的变化。

1. 选择响应和选择差

选择所导致的上下代之间群体均值的变化称为选择响应（R）。选择响应在不同的文献中常被称作遗传进度(Genetic progress)、遗传进展(Genetic advance)、遗传获得或增益 (Genetic gain)、预期进度 (Predicted progress or gain) 等，代表符号有 Rt、GS、G 和 ΔG 等。

假定在选择前，原始群体的平均数为 μ，从原始群体中选择符合目标要求的个体，这些个体组成当选群体，假定当选群体的平均数为 $\overline{x}$。由当选个体交配产生下一代群体(子代群体)，子代群体平均数为 $\overline{y}$。我们把当选群体与子代群体的平均数之差称为选择差(Selection differential)，记为 S，即

$$S = \overline{x} - \overline{y}$$

根据定义，选择响应 R 为

$$R = \overline{y} - \mu$$

2. 选择响应和选择差之间的关系

在不加选择（即全部群体当选）时

$$\overline{x} = \mu = \overline{y}$$

此时将子代均值对其亲代个体进行回归分析，回归方程应通过（μ，μ）点(见图2.2)，即

$$y - \mu = b(x - \mu)$$

现在假定原始群体中有一部分个体当选了，则当选群体的子代均值可根据回归方程预测为

$$\overline{y} = \mu + b(\overline{x} - \mu)$$

即

$$\overline{y} - \mu = b(\overline{x} - \mu)$$

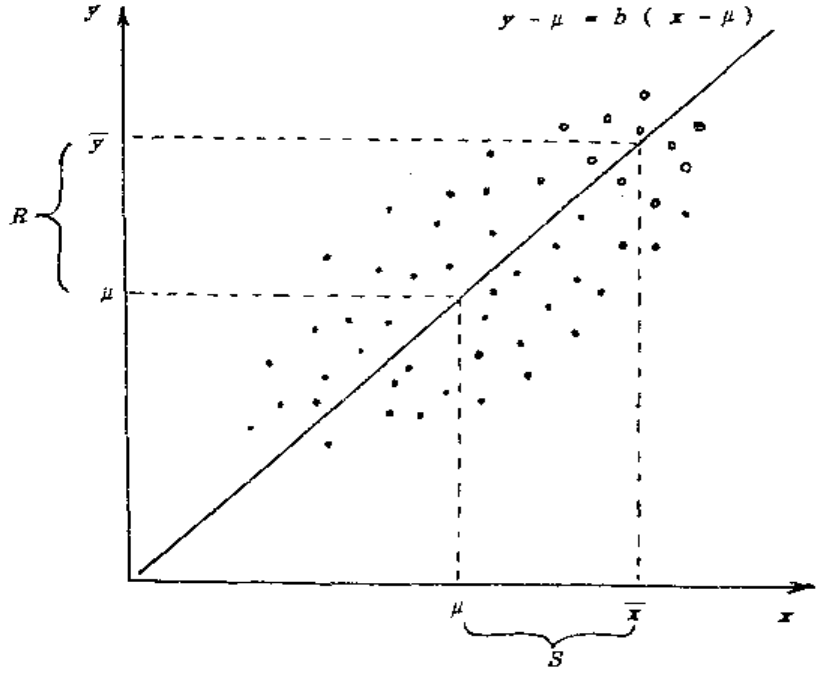


图 2.2 选择对上下代群体均值影响(假定圆圈个体当选)

由于 $R=\overline{y}-\mu$，$S=\overline{x}-\mu$，故

$$R = bS$$

$$R/S = b$$

又因为上下代之间的回归系数类似于遗传率，因此

$$R = h^2S$$

由此可见，选择在下代的实现程度取决于性状的遗传率（图2.3）：

当 $h^2 = 1$ 时，$\overline{y} = \mu + h^2(\overline{x} - \mu) = \overline{x}$，选择完全实现；

当 $h^2 = 0$ 时，$\overline{y} = \mu + h^2(\overline{x} - \mu) = \mu$，选择完全不能实现；

当 $0 < h^2 < 1$ 时，$\overline{y} = \mu + h^2(\overline{x} - \mu)$，选择只能部分实现。

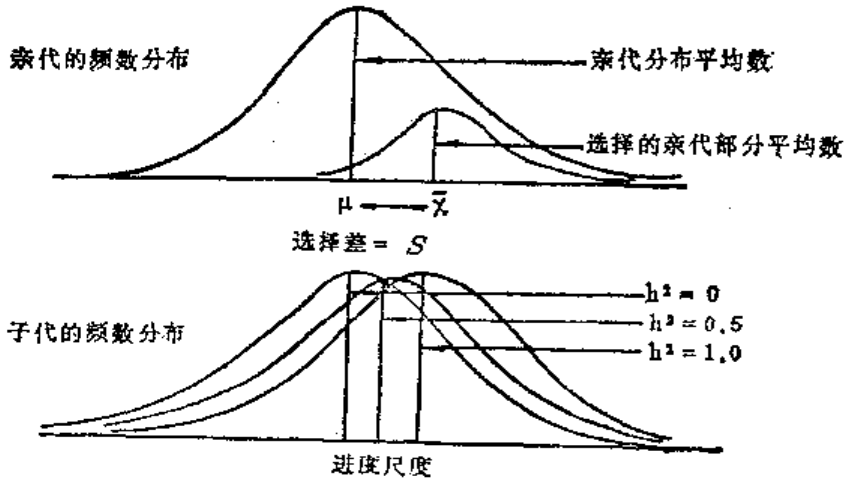


图 2.3 遗传率对选择响应的影响

3. 选择响应的预测与选择强度 (Intensity of selection)

根据公式 $R = h^2S$，要对选择响应进行预测，必需知道性状的遗传率和选择差。而选择差只有在选择结束后才能确定。

影响选择差的因素包括选择率（当选个体占原始群体总个体的比例，p）和选择性状的变异大小（性状的标准差 σ）。当群体变异一定时，选择率越高，选择差就越小；而当选择率一定时，群体变异越大，选择差也越大（图 2.4）。

当选择率一定时，选择差的大小与群体的标准差 σ 成正比，即

$$S = k\sigma$$

其中 k 为比例常数，该常数只与选择率有关，而与 σ 无关。故

$$k = S/\sigma = (\overline{x} - \mu)/\sigma$$

这里 k 为标准化的选择差，称为选择强度。所以

$$R = k\sigma h^2 = k\sigma_g (h^2)^{1/2}$$

其中 σ 实际上为群体性状的表型标准差，而 σ_g 为相应的基因型标准差 $(V_g)^{1/2}$。

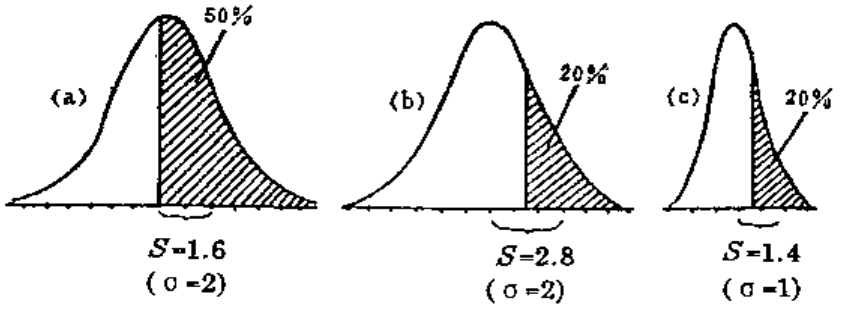


图 2.4 选择率和群体变异程度(性状标准差)对选择差影响

根据上式，提高选择响应的途径包括：(1) 增加群体的遗传变异 σ_g；(2)提高性状的遗传率 h^2；(3) 降低选择率 p，提高选择强度 k。

Simmonds (1979) 给出了求 k 值的经验公式：

$$k = 1.13 + 0.73 \log (1/p)$$

此式要求原始群体样本容量 $n>20$，且 $p=0.1\sim0.001$。

由于 $R/S = h^2$，可根据选择响应的结果反过来估计原始群体的遗传率：

$$h^2 = R/S$$

该遗传率被称为现实遗传率 (Realized heritability)

2.5.5 遗传相关 (Genetic correlation) 与间接选择 (Indirect selection)

1. 定义

统计学中关于相关系数的定义为

$$r_{xy} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n\sigma_x\sigma_y} = \frac{\mathrm{Cov}(x, y)}{\sigma_x \sigma_y}$$

其中Cov(x,y)为 x、y 两个变量的协方差，σ_x 和 σ_y 分别为 x、y 的标准差。在数量遗传学中可将表型相关分解为与遗传和环境有关的两个部分。与方差的分解相似，首先可将表型协方差分解为基因型协方差和环境协方差：

$$\mathrm{Cov}_p = \mathrm{Cov}_g + \mathrm{Cov}_e$$

因而可得 x 和 y 两个性状间的表型相关系数 r_{pxy}，基因型(遗传)相关系数 r_{gxy} 和环境相关系数 r_{exy}：

$$r_{pxy} = \frac{\mathrm{Cov}_p(x,y)}{(\sigma^2_{px}\times\sigma^2_{py})^{1/2}}$$

$$r_{gxy} = \frac{\mathrm{Cov}_g(x,y)}{(\sigma^2_{gx}\times\sigma^2_{gy})^{1/2}}$$

$$r_{exy} = \frac{\mathrm{Cov}_e(x,y)}{(\sigma^2_{ex}\times\sigma^2_{ey})^{1/2}}$$

2. 基因型相关的遗传基础

造成基因型相关的可能原因如下：

(1) 同一基因对不同性状具有多效性，即一因多效；

(2) 控制不同性状的基因存在于同一染色体上，表现连锁遗传；

(3) 控制不同性状的基因对环境具有相同的反应方式；

(4) 同一基因扮演双重角色，既是控制某一性状的主基因，又作为多基因的成员之一参与另一性状的控制。

3. r_g、r_p、r_e 三者之关系

$$r_{pxy} = \frac{\mathrm{Cov}_p(x,y)}{(\sigma^2_{px} \times \sigma^2_{py})^{1/2}} = \frac{\mathrm{Cov}_g(x,y) + \mathrm{Cov}_e(x,y)}{(\sigma^2_{px} \times \sigma^2_{py})^{1/2}}$$

$$= \frac{\mathrm{Cov}_p(x,y)}{(\sigma^2_{px} \times \sigma^2_{py})^{1/2}} \times \frac{(\sigma^2_{gx} \times \sigma^2_{gy})^{1/2}}{(\sigma^2_{gx} \times \sigma^2_{gy})^{1/2}} + \frac{\mathrm{Cov}_e(x,y)}{(\sigma^2_{px} \times \sigma^2_{py})^{1/2}} \times \frac{(\sigma^2_{ex} \times \sigma^2_{ey})^{1/2}}{(\sigma^2_{ex} \times \sigma^2_{ey})^{1/2}}$$

$$= r_{gxy} \times \left[\frac{\sigma^2_{gx}}{\sigma^2_{px}} \times \frac{\sigma^2_{gy}}{\sigma^2_{py}} \right]^{1/2} + r_{exy} \times \left[\frac{\sigma^2_{px} - \sigma^2_{gx}}{\sigma^2_{px}} \times \frac{\sigma^2_{py} - \sigma^2_{gy}}{\sigma^2_{py}} \right]^{1/2}$$

$$= r_{gxy} (h_x^2 \times h_y^2)^{1/2} + r_{exy} \left[(1-h_x^2)(1-h_y^2) \right]^{1/2}$$

可见，在表型相关中，基因型相关占多大比重取决于遗传率的大小：

当 $h_x^2 = h_y^2 = 1$ 时，$r_{pxy} = r_{gxy}$

$h_x^2 = h_y^2 = 0$ 时，$r_{pxy} = r_{exy}$

有时甚至 r_{gxy} 的符号可能与 r_{pxy} 相反。

2.5.6 选择指数 (Selection index) 与综合选择

对多个数量性状进行综合选择的方法包括：(1) 逐项选择法 (Tandem selection)：先对其中一个性状进行选择，当这个性状达到要求后，在保持该性状不变的前提下再进行第二个性状的选择，直到全部目标性状达到要求时为止。(2) 独立淘汰法 (Independent culling levels)：对每个目标性状设立各自的淘汰标准，只要其中一个性状达不到要求即被淘汰。(3) 选择指数法：由与目标性状有关的性状构成一个选择指数，根据选择指数的大小进行选择。下面简单介绍选择指数法。

要选择 y 性状，可根据 y 的相关性状进行选择。为此，必须由各相关性状的表型值 x_1、x_2、…、x_n 构成选择指数 I，使 I 与各相关性状的表型值之间存在线性关系：

$$I = b_1 x_1 + b_2 x_2 + \ldots + b_n x_n$$

其中 b_1、b_2、…、b_n 为各相关性状的加权系数 (经济权重 Weight)。加权系数的大小采用下式估计：

$$\begin{aligned}
&Var_{11}b_1 + Var_{12}b_2 + \cdots + Var_{1n}b_n = g_{1y}\\
&Var_{21}b_1 + Var_{22}b_2 + \cdots + Var_{2n}b_n = g_{2y}\\
&\cdots \qquad\quad \cdots \qquad\quad \cdots\\
&Var_{n1}b_1 + Var_{n2}b_2 + \cdots + Var_{nn}b_n = g_{ny}
\end{aligned}$$

其中 Var_{ii} 为相关性状 i 的表型方差；$Var_{ij} = Var_{ji}$ $(i \neq j)$ 为相关性状 i 和 j 之间的表型协方差；g_{iy} 为相关性状 i 与目标性状（y）的基因型协方差。解上述方程组即可求出 b_i，并由各个体的表型值计算出对应的选择指数 I。

实例 水稻某杂交组合单株穗重(y，目标性状）与每穗粒数（x_1）、二百粒重（x_2）及单株穗数(x_3）的遗传相关显著，故可用 x_1、x_2、x_3 构造 y 的选择指数。

各相关性状的表型方差为：$Var_{11} = 372.02$　$Var_{22} = 0.23$　$Var_{33} = 1.46$

相关性状间的表型协方差为：$Var_{12} = -3.93$　$Var_{13} = -10.98$　$Var_{23} = 0.099$

目标性状与相关性状间的基因型协方差为：$g_{1y} = 24.34$　$g_{2y} = -0.36$　$g_{3y} = 0.58$

根据上述数据可建立以下正规方程组：

$$\begin{aligned}
&372.02\ b_1 + (-3.93)\ b_2 + (-10.98)\ b_3 = 24.34\\
&(-3.93)\ b_1 + 0.23\ b_2 + 0.099\ b_3 = -0.36\\
&(-10.98)\ b_1 + 0.99\ b_2 + 1.46\ b_3 = 0.58
\end{aligned}$$

解得：$b_1 = 0.09$　$b_2 = -0.45$　$b_3 = 1.14$

所以　$I = 0.09\ x_1 - 0.45\ x_2 + 1.14\ x_3$

若有两个株系各相关性状的表型值如下，据此可以计算出各自的选择指数 I。

株系号	x_1	x_2	x_3	I
1	113.3	5.1	8.2	17.3
2	90.7	5.5	8.1	14.9

根据选择指数的大小，株系 1 的丰产性优于株系 2，故株系1当选。

2.6 基因型与环境互作

2.6.1 基因型与环境互作的表现

生物的表现型是基因型和环境共同作用的产物。基因型与环境互作的结果，同一基因型(或品种)在不同环境下有不同的表现。例如A、B两个品种在X、Y两种环境下的表现如图2.5。当存在互作时，两个品种对不同环境的反应是不一致的。(1)虽然在两种环境下A品种的表现均优于B品种，但A品种在Y地表现较佳，而B品种在X地表现较佳（图2.5(1),(2)）。(2) A、B两个品种在X地表现相近，而在Y地表现不一致，反之亦然（图2.5(3),(4)）。(3)两品种在 X、Y两个地点的表现优劣完全相反，在X地，B优于A，而在Y地，A优于B，反之亦然（图5.3 (5)，(6)）。

外界环境的变异会对生物的表型产生影响。环境变异可分为两类：可预测的环境变异

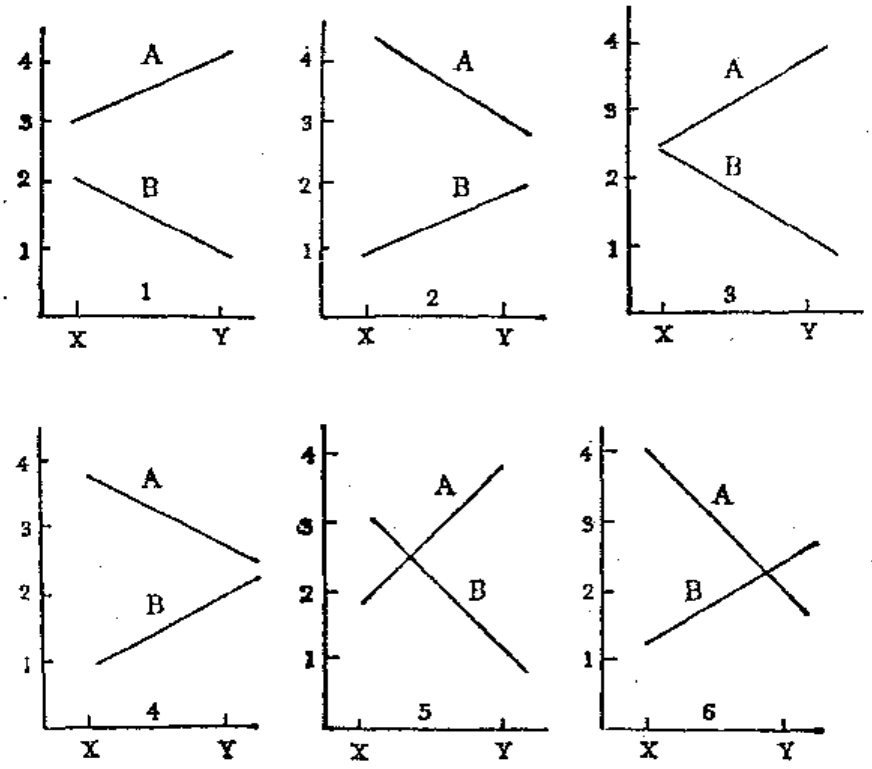


图 2.5　A、B 两品种在 X、Y 两地的表现型（示基因型×环境互作）

（永久性的环境因素，如气候变化，土壤类型，日照长度等）；不可预测的环境变异（天气的变动，日温的变化，病虫害的侵袭等）。

2.6.2　基因型与环境互作的遗传模型

当基因型的效应不受环境变化的影响时，生物个体 i 在 j 环境下表现型的线性公式为：

$$y_{ij} = \mu + g_i + e_j$$

其中 μ 为群体平均数，g_i 为第 i 种基因型的效应，e_j 为第 j 环境的效应。

当基因型和环境之间存在某种互作时

$$y_{ij} = \mu + g_i + e_j + I_{ij}$$

其中 I_{ij} 是第 i 基因型与第 j 环境的互作效应。

根据这一关系，可推导出在 j 环境下纯合亲本及其杂种后代的平均表现型值：

$$\overline{P}_{1j} = m + [d] + e_j + I_{dj}$$

$$\overline{P}_{2j} = m - [d] + e_j - I_{dj}$$

$$\overline{F}_{1j} = m + [h] + e_j + I_{dj}$$

$$\overline{F}_{2j} = m + 1/2\ [h] + e_j + 1/2\ I_{hj}$$

$$\overline{B}_{1j} = m + 1/2\ [d] + 1/2\ [h] + e_j + 1/2\ I_{dj} + 1/2\ I_{hj}$$

$$\overline{B}_{2j} = m - 1/2\ [d] + 1/2\ [h] + e_j - 1/2\ I_{dj} + 1/2\ I_{hj}$$

其中 m 为两亲本在所有环境下的平均值，I_{dj} 为基因加性效应与第 j 环境的互作效应，I_{hj} 为基因显性效应与第 j 环境的互作效应。这些模型的限定条件是

$$\Sigma\ e_j = \Sigma\ I_{aj} = \Sigma\ I_{hj} = 0$$

因此，根据 6 个 世代在不同环境下的表现，不仅可以估算基因效应之值，还可估算环境效应值及互作效应值。

2.6.3. 基因型与环境互作效应的估算

Bucio Alanis (1966) 利用两个烟草纯系品种及其 F_1 的 16 年重复试验资料，估算了上述遗传模型中烟草株高的各个参数。估计式如下：

$$m = 1/n\ \Sigma\ (\overline{P}_{1j} + \overline{P}_{2j})/2$$

$$[d] = 1/n\ \Sigma\ (\overline{P}_{1j} - \overline{P}_{2j})/2$$

$$[h] = 1/n\ \Sigma\ [\ \overline{F}_{1j} - (\overline{P}_{1j} + \overline{P}_{2j})/2\ ]$$

$$e_j = 1/2\ (\overline{P}_{1j} + \overline{P}_{2j}) - m$$

$$I_{aj} = 1/2\ (\overline{P}_{1j} - \overline{P}_{2j}) - [d]$$

$$I_{hj} = \overline{F}_{1j} - 1/2\ (\overline{P}_{1j} + \overline{P}_{2j}) - [h]$$

所得结果列于表 2.16。

结果表明，在大多数情况下，环境效应值 $\hat{e}_j$ 大于基因型与环境互作的的效应值 ($\hat{I}_{aj}$ 和 $\hat{I}_{hj}$)，加性效应与环境的互作 $\hat{I}_{aj}$ 大于显性效应与环境的互作 $\hat{I}_{hj}$。并且，$\hat{e}_j$、$\hat{I}_{aj}$ 和 $\hat{I}_{hj}$ 的大小和方向存在一定的相关性。互作效应值和环境效应值的回

表2.16 两个烟草品种及其 F_1 16 年株高平均数的遗传分析结果

年 份	$\hat{e}_j$	$\hat{I}_{aj}$	$\hat{I}_{hj}$
1	0.08	5.03	0.05
2	1.70	6.04	1.30
3	-0.96	3.68	-0.38
4	-12.83	-2.11	-3.30
5	-6.68	-3.86	-3.53
6	-16.10	-8.56	-1.78
7	-12.34	-4.29	-8.28
8	-2.31	-3.18	-2.54
9	-4.75	-6.91	-1.75
10	-25.02	-11.13	-6.12
11	4.42	-3.56	2.46
12	-8.97	2.64	1.47
13	26.72	7.92	8.69
14	21.08	5.84	5.38
15	13.11	3.15	1.17
16	22.84	10.03	2.08

$\hat{m} = 112.27$　　$[\hat{d}] = 7.34$　　$[\hat{h}] = 13.94$

归分析表明，I_{aj} 与 e_j 的回归系数为 $b_a = 0.352 \pm 0.066$，I_{bj} 与 e_j 的回归系数为 $b_b = 0.236 \pm 0.043$。根据这一回归关系，可将世代均值的表达式改写为：

$$P_{1j} = m + [d] + (1+b_a)\ e_j$$
$$P_{2j} = m - [d] + (1-b_a)\ e_j$$
$$F_{1j} = m + [h] + (1+b_h)\ e_j$$

据此可将世代均值分解为两部分：(1) 平均表现 —— 群体在各环境中的平均表现，它取决于 m、$[d]$、$[h]$。(2) 环境敏感性 ——取决于环境效应值 e_j 和群体平均数对 e_j 的回归系数 $(1+b_a)$、$(1-b_a)$ 和 $(1+b_h)$。

采用多个品种（基因型）在多个环境下进行的重复试验（例如品种区域试验），可应用方差分析或回归分析的方法，分析基因型×环境互作效应。详细分析方法见Lin和Binns (1986) 和 Westcott (1986)。

2.6.4 基因型与环境互作分析在植物育种中的应用

1. 品种稳定性的机制及测定

品种稳定性的三种类型：(1)当基因型在各环境间的方差较小时是稳定的；(3) 当基因型与环境的反应与试验中所有基因型的反应是平行的时候，认为是稳定的；(3) 当基因型与环境指数的回归模型的剩余均方较小时，认为是稳定的。

2. 异地选择效果的评价

如果在异地 (X) 对性状进行选择，希望入选后代在当地 (Y) 仍有较好的表现，性状在 Y 的选择响应为

$$R = k\ (h_X^2)^{1/2}(h_Y^2)^{1/2} r_{g\,XY}\ \sigma_{P(Y)}$$

由此可见，性状在两地所表现的遗传相关越高，异地选择的效果就越好。

实际应用时通常考虑异地选择的相对效果，即相关选择响应 R_{XY} 与直接在当地选择响应 R_Y 的比值：

$$\frac{R_{XY}}{R_Y} = r_{g\,XY}\frac{k_X(h_X^2)^{1/2}}{k_Y(h_Y^2)^{1/2}}$$

其中 $R_{XY} = k_X(h_X^2)^{1/2}(h_Y^2)^{1/2} r_{g\,XY}\ \sigma_{P(Y)}$

$$R_Y = k_Y h_Y^2\ \sigma_{P(Y)}$$

当两地选择强度相等时，相对选择效果为

$$\frac{R_{XY}}{R_Y} = r_{g\,XY}\frac{(h_X^2)^{1/2}}{(h_Y^2)^{1/2}}$$

这表明异地选择的相对效果取决于在两地表现型的遗传相关性以及两地遗传率的比值。如果遗传相关很低，或者性状在异地的遗传率很低，则异地选择的效果就很差，应强调在当地直接选择。

第 3 章　遗传标记

在遗传学实验中通常将可识别的等位基因称为遗传标记，并用来研究基因遗传和变异的规律。等位基因是在染色体上占据同一座位的基因的不同形式，也是在生物群体中遗传多态性的表现形式。基因组DNA的变异是遗传多态性形成的分子基础。在真核生物中，基因组DNA多态性要远远多于各种基因的遗传多态性，因为在基因组中存在着大量的不编码的DNA序列，它们的变异不受或较少地受自然选择的制约。然而由于直接检测DNA多态性的方法学的限制，将DNA多态性用作遗传标记还是1980年以后的事情。

遗传标记经常用在两类遗传学实验中。第一类是遗传学研究中最基本的工作，即连锁分析、基因定位和遗传作图。这也是本书讨论的重点。第二类实验是研究基因的转移。将遗传标记作为其所在染色体、染色体区段、座位及外源DNA片段的标记。在传统的杂交育种过程中常将标记基因作为供体亲本的染色体及其片段掺入的判断依据，而在细胞融合及基因工程研究中，遗传标记的应用已成为选择和筛选的主要手段。追溯遗传学的历史不难发现，由突变形成的多个等位基因的发现，是遗传学研究的生长点。摩尔根在遗传学上的全部贡献几乎都可以归功于他对果蝇遗传标记的开拓和应用。早在1908年，摩尔根在哥伦比亚大学的实验室已开始饲养果蝇。1910年5月，一只奇特的果蝇被发现，它的眼色不是野生型的红色而是白色。正是这只白色果蝇成为近代遗传学研究的起点。摩尔根用白眼雄蝇与野生型杂交，证明了在F_1代中白眼为隐性性状，在F_2代中也验证了红眼与白眼的遗传服从孟德尔定律，以3:1分离。但是在按雌雄性别分别记数时，发现了异常现象。F_2的雌蝇全部为红眼，雄蝇中一半为红眼一半为白眼，雌雄之比为正常的1:1。从这一复杂的遗传现象中，摩尔根悟出了一个规律，即决定眼色的基因(R)与决定性别的基因(X)是结合在一起传递的。白眼标记的发现极大地激发了摩尔根实验室发现各种遗传标记的兴趣。到1912年底，共发现了40多个肉眼可辨的果蝇突变型。他们每得到一个突变型就立即作杂交实验。由于果蝇十天即可繁殖一代，研究进展非常迅速。1913年，A. H. Sturtevant将果蝇的5个基因经连锁分析定位在X染色体上，从此确定了遗传学的染色体理论和遗传作图的基本原理。这些原理适用于从噬菌体、细菌一直到高等动植物乃至人类的一切生物，而且一直沿用至今。

然而，在高等动植物的遗传作图工作中，大多数科学家远远不如摩尔根那么幸运，原因不仅仅在于这些生物基因组的复杂性或者它们的有性世代太长，主要的瓶颈因素是可供应用的遗传标记数有限。许多农作物和家畜的遗传图谱是集几代遗传学家和育种学家辛勤劳动的结晶。到八十年代初，已建成了一百多种生物的遗传图谱，其中如玉米和番茄的遗传标记已达数百个。这些图谱的构建，也可以说是分子生物学家未曾参与的早期的基因组研究，因为用于作图的遗传标记都是以等位基因的表型为识别依据的。应该说在果蝇的X染色体连锁群初建以后长达六十年的时间里，与遗传学主流的突飞猛进相比，遗传标记研究开发的节奏是极其缓慢的。遗传标记的研究达到分子水平比分子遗传学迟到了二十多年。

当然，在这个漫长的时期里，遗传标记的范围也有相当的拓展。随着生物学各个分支学科的发展，遗传标记也从外观形态的表现型，扩展到生理、生化、细胞、发育、病理和免疫等多个方面。但是，所有这些不同类型的标记都是以生物学的性状形式进行判别的，即在检测时几乎离不开生物的活体形式。如细菌的抗药性基因是以人工培养基上菌落的形成与否为标记的，植物的抗病基因是从叶片或植株的其它部分的病斑有无或多寡、大小等为标记的。即使在生化遗传学中分析参与代谢途径的酶的基因时使用的营养缺陷型标记，仍然要依靠在特定培养基上菌落的形成与否进行识别。直到五十年代和六十年代之交，同工酶标记的兴起，遗传标记的识别才突破了活体的形式。所谓同工酶标记是指编码功能相同、结构和组成略有不同的一类酶的等位基因。同工酶的不同形式可以通过电泳或色谱的方法予以区分，并用专门的染色反应予以显示。从检测方法的角度看，同工酶标记是一类新型的分子标记，但是仍然没有突破表达基因的范围。由于方法和技术上的限制，迄今经常使用的同工酶标记的数目还是相当有限。

美国的Botstein等(1980)首先提出DNA限制性片段长度多态性(RFLP)可以作为遗传标记，从此开创了直接应用DNA多态性发展遗传标记的新阶段。从理论上讲，DNA测序是揭示DNA多态性最彻底最全面的方法，但在实践中很难应用。限制性内切酶可以识别DNA分子上的靶序列(通常为4～6个碱基对)，任何靶序列的点突变或2个相邻靶序位间的DNA序列的插入或缺失，都会改变相应的DNA片段的长度，成为电泳分离和分子杂交检测的基础。RFLP标记的应用实现了两个突破，一是遗传标记的数量不再局限于基因的遗传多态性，二是遗传标记的检测不受生物活体的表型限制。RFLP标记的诞生大大加速了各种生物遗传图谱的建立和发展，同时也提高了基因定位的精度和速度。而且使遗传图谱的饱和成为可能，这也从一个方面推动了人类基因组和各种基因组计划的出台和进展。

八十年代后期，DNA多聚酶链式反应(PCR)的发展，也使直接扩增DNA的多态性成为可能，在此基础上又产生了各种新型的分子标记，如RAPD标记和SSR标记等。与此同时，各种电泳技术的发展，也提高了电泳分辨DNA多态性的能力，如分辨单链DNA构象的电泳技术产生了SSCP标记。由于各种分子标记的应用，遗传学又进入了一个日新月异的发展阶段。下面将简要介绍目前常用于遗传作图的各种遗传标记，重点讨论分子标记。

3．1 形态标记

遗传学研究中所采用的遗传标记是随着遗传学，特别是基因概念的发展而发展的。本世纪初，孟德尔以豌豆为材料，详细地研究了七对相对性状的遗传，由于这些性状都具有典型的外部形态特征，很容易识别，因而使得孟德尔可以对杂种后代的不同个体的性状进行分组和归类，从而超越他所处的时代，提出了生物性状遗传的“遗传因子”假说和两个遗传定律——分离规律和独立分配规律。正是由于形态性状在外部特征上所具有的鲜明性而使之可以用作一种遗传标记来研究性状间的相互关系。因此，形态标记就是一种特定的肉眼可见的外部特征，而且影响这一特征的基因及其所在的染色体、基因与相邻基因在染色体上的相对位置都已得到明确阐明，可以用其标记某一染色体区段，并对其它未定位的基因进行连锁分析。从广义上讲，形态标记还包括与色素、生理特性、生殖特性、抗病

虫性等有关的标记。

形态性状基因的染色体定位最初就是通过性状间的关联和经典的二点、三点测验进行的。通过判断不同性状间的遗传是否偏离依照独立分配所预期比例来确定控制这些性状的基因是否位于同一染色体上。采用这种方法把相互连锁在一起的基因定为一个连锁群，并与假定的物质实体——染色体相对应，以性状间重组率的大小作为这些基因在染色体上的相对距离，并确定其毗邻关系。1906年，W. Bateson 等在香豌豆杂交试验中根据花色与花粉粒形态两对性状间的相互关联，摩尔根根据对果蝇六个形态性状的分析发现了连锁遗传现象。采用这种方法，确定了生物中最早发现的基因连锁群，并从中选择其形态特征明显，易于识别的基因作为形态标记，并创了利用已知染色体相对位置的遗传标记定位未知基因的先例。摩尔根领导的实验室根据研究性状与形态标记的连锁关系，建立了完整的果蝇连锁群。1928年，赵莲芳 (1928) 根据8个水稻品种的杂交资料，研究了茎、叶和颖花等器官的颜色、护颖长度、谷粒外形等12种性状的遗传，由其中9个基因的相互关系，确定了水稻的三个连锁群。借助于这种方法在一些栽培物种中确定了许多质量性状基因的连锁关系及其在染色体上的相对位置，并绘制出较为完整的遗传连锁图谱。以形态标记为基础的连锁群的建立为生理、生化性状的遗传研究奠定了基础。

带有形态标记性状的遗传材料被称为标记基因材料，它们是遗传学研究的宝贵财富。标记基因材料的收集、保存和利用历来受到各国研究者的重视。例如在水稻中，国际水稻所和日本系统地收集了大量的标记基因材料，并作为重要的种质资源加以保存。据Khush (1987) 报道，保存于国际水稻遗传资源中心的各类标记基因材料共255份，国际水稻遗传委员会报道了目前国际上已公认的300多个水稻标记基因材料，其中大部分带有形态学标记性状。表3.1 给出了水稻的部分标记性状的名称及其已鉴定的部分基因。这些形态性状大多也存在于不同的植物种中，尤其是禾谷类作物中。类似地，同属一类的哺乳类动物，用于描述其形态特征的性状也是大同小异的。

由自然突变或物理化学诱变均可获得具有特定形态特征的遗传标记材料。例如在组织培养及诱变育种过程中经常出现大量的变异材料，其中不乏在形态特征或生理特性上具有特殊表现的个体，经过选择，就可获得稳定遗传的形态标记材料。然后，利用已知基因座位的标记材料与之杂交，通过连锁遗传分析就可能将其变异特性定位到染色体上，这个变异材料就成为一种新的标记基因材料。在水稻中，目前已有的标记基因材料，大多是经过精心选育获得的。有的仅带有一个标记基因，有的则是经过不同标记基因材料的相互杂交和选择而获得的带有多个标记基因的材料。后者用于基因连锁分析时，可同时分析几个标记性状。

3.2 细胞学标记

1910年以后，摩尔根将孟德尔所称的“遗传因子”的行为与细胞核染色体的行动结合起来进行研究，将这些“因子”与染色体上呈直线排列而染色明显的念珠状颗粒物对应起来，从而将“遗传因子”—— 基因落实为染色体上占有一定位置的实体。这导致了细胞遗传学的诞生。通过对不同物种染色体形态、数目和结构的研究，偶尔会发现某些个体比

表 3.1　水稻的部分形态学标记性状及其基因(根据Rice Genet. News. Vol.6)

性状分类	标记性状	已鉴定的基因举例
色素	紫色叶耳	*Pau*
	紫色叶舌	*Plg*
	紫色谷壳	*Pr*
	黑色谷壳	*Bh*
	白色谷壳	*Wh*
	紫色柱头	*Ps-1, Ps-2, Ps-3*
	紫色叶鞘	*Psh*
	谷壳棕色条纹	*Bf*
叶绿素异常	白化苗	*al-1, al-2, al-3, ...*
	缺绿	*chl-1, chl-2, chl-3, ...*
	黄化苗	*v-1, v-2, v-3, v-4, ...*
	斑马条纹叶	*z-1, z-2, z-3, z-4, ...*
矮生	小粒矮生	*d-1,d-7, d-11, d-13, ...*
	多蘖矮生	*d-3,d-4, d-5, d-10, ...*
	非正常矮生	*d-2, d-6, d-20, d-21, ...*
	半矮生	*d-9, d-12, sd-1, sd-2, ...*
籽粒	有芒	*An-1, An-2, An-3, An-4*
	内颖发育不全	*dp-1, dp-2*
	大胚	*ge*
	圆粒	*rk-1, rk-2*
	酚染色反应	*Ph*
	小粒	*Mi*
	谷壳叶化型不育	*oms*
	多个雌蕊	*mp-1, mp-2*
胚乳	暗色胚乳	*du-1, du-2, du-3, du-4, ...*
	胚乳贮藏蛋白	*esp-1, esp-2, esp-3*
	胚乳皱缩	*shr-1, shr-2*
	糯性胚乳	*wx*
	粉质胚乳	*flo-1, flo-2*
穗	密穗	*Dn-1, Dn-2, dn-3*
	穗颈叶	*nl-1, nl-2*
	簇生小穗	*Cl*

表 3.1　水稻的部分形态学标记性状及其基因（续）

性状分类	标记性状	已鉴定的基因举例
茎	脆性茎秆	*bc-1, bc-2, bc-3, ...*
	直立生长	*er*
	匍匐生长	*la*
	细茎	*fc-1, fc-2*
	最上节间伸长	*eui*
叶	无叶耳	*aul*
	无叶舌	*lg*
	光叶	*gl-1, gl-2*
	窄叶	*nal-1, nal-2, nal-3, nal-4, ...*
	卷叶	*rl-1, rl-2, rl-3, rl-4, ...*
	褐色叶斑	*bl-1, bl-2, bl-3, bl-4, ...*
	斑点叶	*spl-1, spl-2, spl-3, spl-4, ...*
根	根生长受抑制	*rt*
抽穗期	早熟	*Ef-1*
	迟开花	*lf-1, lf-2*
	感光性	*Se-1, se-2*
育性	雄性不育	*ms-1, ms-7, ms-8, ms-9, ...*
	雄性部分不育	*ms-2, ms-3, ms-4, ms-5, ...*
	配子体F_1不育	*s-a-1, s-c-1, s-d-1, ...*
	互作不育	*S-1, S-2, S-3, S-4, ...*
	细胞质雄性不育	*cms-bo, cms-WA, cms-ld, ...*
	育性恢复	*Rf-1, Rf-2, ...*
配子体基因		*ga-1, ga-2, ga-3, ...*
抗病性	稻瘟病抗性	*Pi-a, Pi-b, Pi-f, Pi-i, ...*
	白叶枯病抗性	*Xa-1, Xa-2, Xa-3, Xa-4, ...*
	草丛矮缩病抗性	*Gsv*
抗虫性	褐稻虱抗性	*Bph-1, bph-2, Bph-3, ...*
	白背飞虱抗性	*Wph-1, Wph-2, Wph-3, ...*
耐逆性	苗期耐冷性	*Cts-1*
	敦稗敏感性	*prs*

该物种的正常合子染色体数增加或减少了一个以至若干个染色体，这种个体称为非整倍体。缺体、单体、三体、端着丝点染色体等非整倍体，缺失、易位等染色体结构变异以及各种异形染色体（如染色质超数染色体）等都有其特定的细胞学特征，在其与其它具有正常染色体品系之间的杂种或衍生系中，由于减数分裂过程的异常，容易导致特定染色体上基因的分离与重组发生偏离，因而可以作为一种遗传标记来测定基因所在的染色体及其相对位置或通过染色体代换等遗传操纵进行基因定位。在玉米中利用B－A易位系，水稻中利用易位系，小麦中利用端着丝点染色体，大麦中利用端着丝点三体，棉花中利用易位和单体等

非整倍体，在番茄中利用各种三体，在烟草中利用单体，已成功地将许多质量基因定位于染色体上。并建立了相应的连锁群。

在小麦中，非整倍体已广泛应用于定位基因所在的染色体，其中单体(2n=41) 在这类分析中特别有效。同时由于单体和端着丝点染色体出现在小麦的每一对染色体中，因而已产生出完整的品种间21个染色体的代换系，使每一品种的每一对染色体分别被一共同亲本的对应染色体所取代。这样，两个品种间的遗传差异可以一个一个染色体的分割开来，而且每个代换系就象一个品种一样成为一个稳定的纯系，可以用于重复试验。在小麦中利用端着丝点染色体还成功地确定了基因所在的染色体臂位。

染色体结构变异体和非整倍体等细胞学标记材料，常常具有特定的形态学特征，使其很容易与对应的二倍体区别开来。表3.2 给出了水稻从籼稻品种 IR36所衍生的十二条染色体的初级三体名称及其形态特征。根据这些特征，在大多数情况下可以将水稻初级三体与其二倍体的IR36 区别开来。此外，培育带有形态标记性状的非整倍体及染色体结构

表 3.2　水稻IR36初级三体的形态特征(Khush等 1984)

染色体*	三体名称	特　　征
1	三体 1（草状）	生长缓慢，穗高度不育，籽粒较细长，稍呈三角形，叶窄，与草相似
2	三体 2（矮生）	小穗短，护颖较长，自交高度不育，叶片短且基部附近常扭曲
3	三体 4（不育）	植株最矮，叶片短厚呈革质，中脉凸出，穗伸出不完全，自交完全不育
4	三体12（高秆）	植株最高，叶片下垂，叶舌长，完全可育
5	三体 5（叶片扭曲）	叶片短且扭曲，有密生茸毛，穗短而紧密，结实率高
6	三体 3（有芒）	籽粒有芒，叶片厚而半卷，叶舌长，自交高度不育
7	三体 7（窄叶）	叶片窄而半卷，穗稍疏松，不完全伸出，部分可育
8	三体 8（卷叶）	叶片窄卷，叶舌短，籽粒短而粗，部分可育
9	三体 9（粗状）	叶片厚而浓绿，茎秆粗，小穗最大，百粒重最高
10	三体10（短粒）	叶舌有毛，叶片直立，完全可育
11	三体11（拟正常）	形态与二倍体姊妹系无法区别，完全可育，需作细胞学鉴定
12	三体 6（丛生）	外型呈丛生草状，穗部顶端小穗退化，育性高

*采用国际水稻遗传协会1990年制订的染色体编号系统

变异衍生系，即将形态标记与细胞学标记结合起来，也可以在杂种后代中直接根据形态标记确定个体的细胞学特征，而不必进行染色体鉴定。以上这些都有助于提高细胞学标记的利用效率。

3.3 生化(同工酶)标记

1941年美国遗传学家G. W. Beadle和生化学家E. L. Tatum 通过研究红色面包霉的生化突变型，对一系列营养缺陷型的遗传分析提出了“一个基因一个酶”的假说，创立了生化遗传学。50年代许多科学家发现同一种酶可具有多种不同的形式。同时由于淀粉凝胶技术的发展和组织化学染色剂的使用使这种酶的多种形式成为肉眼可辨的带型——酶谱。Market和Moller(1959)根据对几种动物乳糖脱氢酶 *LDH* 的多种形式的研究提出 isozyme 一词来描述具有同一底物专一性的不同分子形式的酶，并证实了同工酶具有组织、发育及物种的特异性。

和大多数其它基因一样，编码酶的基因也可以通过适当的方式定位于染色体上。在具有大量已经定位形态标记的物种中，可以通过二点或三点测验对同工酶基因进行定位。在没有已定位的标记可以利用时，可根据同工酶基因座位等位基因的共显性特性，通过产生多个酶座位处于杂合状态的F_1，可以同时确定这些同工酶座位的连锁关系。随着已定位的同工酶座位数的增加，就可以利用它们作为一种遗传标记来迅速确定未定位的其它基因。这种同工酶标记是建立在生化遗传学基础上的，所以又称为生化标记或蛋白质标记。有时也被当作一种分子标记。

表3.3 给出了水稻中已定位的一些同工酶基因及其所在的染色体，同一基因座位的不同等位基因数以及带有这些等位基因的标记品系。

3.4 分子标记

这里所指的分子标记是以DNA多态性为基础的遗传标记。

1953年，J. D. Watson和F. H. C. Crick提出了DNA分子结构的双螺旋模型，圆满地解释了DNA就是基因的有机化学实体，宣布了分子遗传学时代的到来。现代基因概念的发展使直接利用DNA分子中核苷酸序列的变异作为遗传标记成为可能。真核生物的遗传信息储存于核染色体和细胞器基因组的DNA序列中。DNA分子很大，比细胞中任何其它分子都要大得多。DNA通常用每个细胞的绝对量(Pg)或碱基对数(千碱基对kb或百万碱基对Mb)表示。由于DNA分子很相似，都是由A-T或G-C核苷酸对组成)，所以很容易采用下面的关系式来进行两种计量单位的换算：

$1\ Pg = 0.965\times10^{9}bp = 6.1\times10^{11}$ dalton

尽管生物能快速、精确地复制自己的 DNA，但有许多机制能引起DNA的变化。可能出现一个碱基的变化，也可能由于倒位、易位、缺失或转座导致多个乃至一长段碱基对的变

表 3.3 水稻的同工酶基因（根据 Rice Genet. Newsl. Vol. 7）

同工酶	基因座位	染色体	等位基因	标记品系
酸性磷酸酯酶	*Acp-1*	12	0～3	NIG1707, IR36, T65 NIG1707
	Acp-2	12	0,1	T65, NIG1707
	Acp-4	7	1,2	R23754, R6245
乙醇脱氢酶	*Adh-1*	11	0～3	NIG624,T65,NIG244,NIG437
氨基肽酶	*Amp-1*	2	0～6	R29040,IR36, R6245,NIG413 R33888,R27748,R9011
	Amp-2	8	1～4	T65,IR36,R6434,R32301
	Amp-3	6	0～6	R33888,T65,NIG244,NIG729 R9158,R33153,R6538
	Amp-4	8	1,2,3	R6274,R33191,R33888
β-淀粉酶	*βAmy-1*	7	0～3	HO246,LO1050,T65,NO1208
过氧化氢酶	*Cat-1*	6	1,2,3	IR36,T65,R43675
内肽酶	*Enp-1*	6	0,1	R9145,R8896
酯酶	*Est-2*	6	0,1,2	T65,NIG244,IR36
	Est-3	9	1,2	Sinaba, Moxiangu
	Est-5	1	0,1,2	R37801,IR36,NIG721
	Est-7	7	0,1	Apura Irat177
	Est-9	7	1,2	T65,IR36
谷氨酸脱氢酶	*Gdh-1*	3	1,2	IR36,NIG259
天门冬氨酸氨基转移酶	*Got-1*	1	1,2,3	R6274,R6245,R42469
	Got-2	6	1,2	藤坂5号, Texas Fortuna
异柠檬酸脱氢酶	*Icd-1*	1	1,2,3	R6274,Apura, R9158
苹果酸脱氢酶	*Mal-1*	1	1,2	T65,IR36
过氧化物酶	*Pox-1*	5	0,1,2	CHU3008,NIG504,CHU8216
	Pox-2	12	0,1	T65,NIG108
	Pox-5	6	1,2	R6247,R6538
	M-Pox-1	5	0,1	Nig108,TAR701
磷酸葡萄糖异构酶	*Pgi-1*	3	0～3	NIG-J328,IR36,NIG504 CNPAF800157
	Pgi-2	6	1～4	T65,IR36,R6245,R12800
磷酸葡萄糖脱氢酶	*Pgd-1*	11	1,2,3	T65,IR36,NIG108
	Pgd-2	6	1,2	T65,NIG624

化。由于细胞的DNA 含量很高，因此不大可能有两个生物DNA的碱基序列是相同的。通常在动植物的自然群体中存在大量的 DNA 变异。有几种方法可用于揭示 DNA序列的自然变

异。一种方法是直接测定DNA 的序列，然后进行仔细的比较，遗憾的是该方法既繁琐又花时间，在遗传学的常规研究中难以应用。

3.4.1 RFLP标记

1980年，Botstein首先提出用RFLP作为标记构建遗传连锁图谱的设想(Botstein 等1980)，并在1987年由Donis-Keller等建成了第一张人的RFLP图谱 (Donis-Keller等1987)。所谓 RFLP，是指用限制性内切酶酶切不同个体基因组DNA 后，含同源序列的酶切片段在长度上的差异。差异的显示和检测是用克隆的DNA 片段作为同源序列探针 (即 RFLP标记) 进行分子杂交，再通过放射自显影 (或非同位素技术) 实现的。

限制性内切酶能识别DNA 顺序上由特定碱基组成的识别位点(即限制性位点)，如果DNA上存在这种碱基序列，限制性内切酶就可以在识别位点切开 DNA。因此，限制性内切酶能将很长的DNA分子降解成许多长短不一的小片段，片段的数目和长度反映了 DNA分子上限制性酶切位点的分布。对于特定的DNA/限制性内切酶组合，所产生的片段是特异的，它能作为某一DNA (或含有该 DNA的生物)的特有“指纹”。对于相对较小的DNA分子，如叶绿体DNA，当用一典型的限制性内切酶(如 *EcoR* I)降解时，通常产生40个左右不同的限制性片段。纯化的叶绿体DNA 酶解后所产生的限制性片段经琼脂糖凝胶电泳，就可根据片段的大小将其分开。凝胶用溴乙锭染色后，可直接在紫外灯下观察或拍照。碱基序列互不相同的、或已发生插入、缺失、易位的叶绿体DNA都会产生不同大小的限制性片段。这种片段大小的差异称为限制性片段长度多态性 (Restriction Fragment Length Polymorphisms, RFLPs)，简称RFLP。它能用于在DNA水平上直接测定遗传变异。

在将RFLP 分析用于真核生物染色体 DNA时，分析过程要复杂得多。由于染色体 DNA分子很长，用典型的限制性内切酶来酶解高等植物的核DNA 会产生数百万个DNA片段，其大小变化是连续的，如把酶解的DNA 进行电泳，用溴乙锭染色后，无法区别单个的片段，因为如此之多的DNA 片段紧密排列为连续的一片。然而，各个限制性片段在凝胶中还是分开的，只是肉眼不能分辨，需要借助于更为复杂的技术程序才能将不同生物或同一生物不同品种的限制性片段的差异区分开来。这些技术程序包括 Southern吸印，DNA 探针的制备，同位素或非同位素标记以及DNA分子杂交。由于核DNA 的RFLP 不能直接观察，通常采用染色体DNA 的小片段作为探针(Probe)通过专一性很高的 DNA-DNA 杂交去检测各个DNA 片段。这些探针能够检测出酶解产生的复杂核DNA片段混合物中与其同源的限制性片段。要采用这一技术，必须事先准备一系列作探针用的染色体DNA 片段。为了获取作探针用的DNA 片段，可从所要研究的生物中提取DNA，经限制性内切酶酶解，产生较小的片段(通常2～5kb)。但是直接分离各个限制性片段非常困难，为此可采用分子克隆技术中建立基因文库的方法，通常把各个限制性片段连接到细菌质粒中，然后用重组的质粒转化细菌细胞。细菌在生长分裂的同时复制了质粒。这样只要分离质粒，就能提供大量的某一生物的某种DNA 限制性片段，用作杂交探针。带有某一片段的菌系能保持很长一段时间，因此，同一探针能反复地使用，或提供给其他的研究者共享。

由于技术本身的限制，只有单拷贝 (或少数拷贝) 的序列才能用作探针，在杂交显影后显示清晰可辨的带型。常用探针的主要来源是随机的基因组(Random genome) 克隆(简

称RG克隆）和cDNA克隆。

1．RG克隆：来源于基因组文库。与通常的基因组文库相比，这类文库的构建比较简单，不要求分离基因所必需的对基因组的高度覆盖率。但为了获得尽可能多的RFLP标记，用于RFLP分析的基因组文库应尽可能多地富集单拷贝序列，为此，可用甲基化敏感型的限制性内切酶（如 *Pst* I）切割基因组DNA，获得长度为1～2kb的片段，再与质粒DNA连接、转化、建库。由于甲基化一般不发生在表达的基因组DNA中，因而所获得的可克隆片段大多来自单拷贝的结构基因区。通常，文库中的每一个单拷贝序列的克隆均可用于RFLP分析，但只有在亲本间显示多态性的克隆，才能作为分子标记。

2．cDNA克隆：来源于cDNA文库。一般来讲，用于RFLP分析的cDNA文库没有特殊的要求，但为了获得基因组表达广泛基因的代表性，不宜采用高度专化的组织器官制备cDNA文库。不过可分别制备不同组织器官的 cDNA 文库，以扩大 cDNA 序列在基因组中的代表性。由于cDNA克隆都是结构基因，因此绝大部分的克隆都是单拷贝的。虽然cDNA 序列本身有相当大的保守性，但由于与cDNA 序列相应的基因内部的内含子序列和基因5′和3′端非编码区序列的变异性，因而用cDNA 作RFLP 分析有可能揭示更大的多态性。

为了利用 DNA-DNA 杂交技术检测特定的DNA 片段，探针DNA 和凝胶上的DNA必须是单链（变性）的。将凝胶浸在碱液 (NaOH) 中使 DNA变性。为了便于杂交，DNA通过 Southern 印迹转移的方法从凝胶转移到硝酸纤维素滤膜之类（目前多采用专门的尼龙滤膜，因为这种滤膜强度大，可多次反复使用）的固相支持物上，使变性的DNA 带紧紧地与膜结合。然后使某一克隆探针变性，在合适的温度和盐浓度下，变性的探针与膜上同源的核DNA限制性片段专一地杂交。 然而，为观察到探针在什么地方杂交，就必须用某种方法对探针进行标记。最常用的方法就是用放射性同位素^{32}P标记探针。让标记的探针与含DNA 的滤膜杂交，滤膜经放射自显影，就可发现探针在何处与限制性片段发生了杂交。 采用这一方法可以检测出复杂的基因组 DNA的酶解液中的每一个限制性片段。整个RFLP 分析过程如图3.1所示。现已研制成功了一种不需要采用放射性同位素而借助于化学发光原理的探针标记技术。由此所揭示的RFLP 可称为 c-RFLP。与此相对应，由放射性同位素标记所揭示的RFLP可称为r-RFLP。

与传统的遗传标记相比，RFLP标记具有下列优点：(1) 无表型效应，RFLP标记的检测不受环境条件和发育阶段的影响； (2) RFLP标记在等位基因之间是共显性的，因此在配制杂交组合时不受杂交方式的影响。 (3) 在非等位的RFLP标记之间不存在上位效应，因而互不干扰。 (4) RFLP标记源于基因组DNA的自身变异，在数量上几乎不受限制。所有这些优点，使得RFLP标记特别适合于构建遗传连锁图谱。表3.4总结了RFLP作为遗传标记的主要特征。

3.4.2 AFLP标记

AFLP 是扩增片段长度多态性 (Amplified Fragments Length Polymorphism) 的简称。相对于RFLP 而言，在AFLP 分析中显示多态性的DNA片段不是由于限制性内切酶酶切基因组DNA产生的，而是通过DNA 多聚酶链式反应 (Polymerase Chain Reaction，PCR) 扩增基因组DNA 的模板产生的。现简单地描述PCR反应的原理与方法。

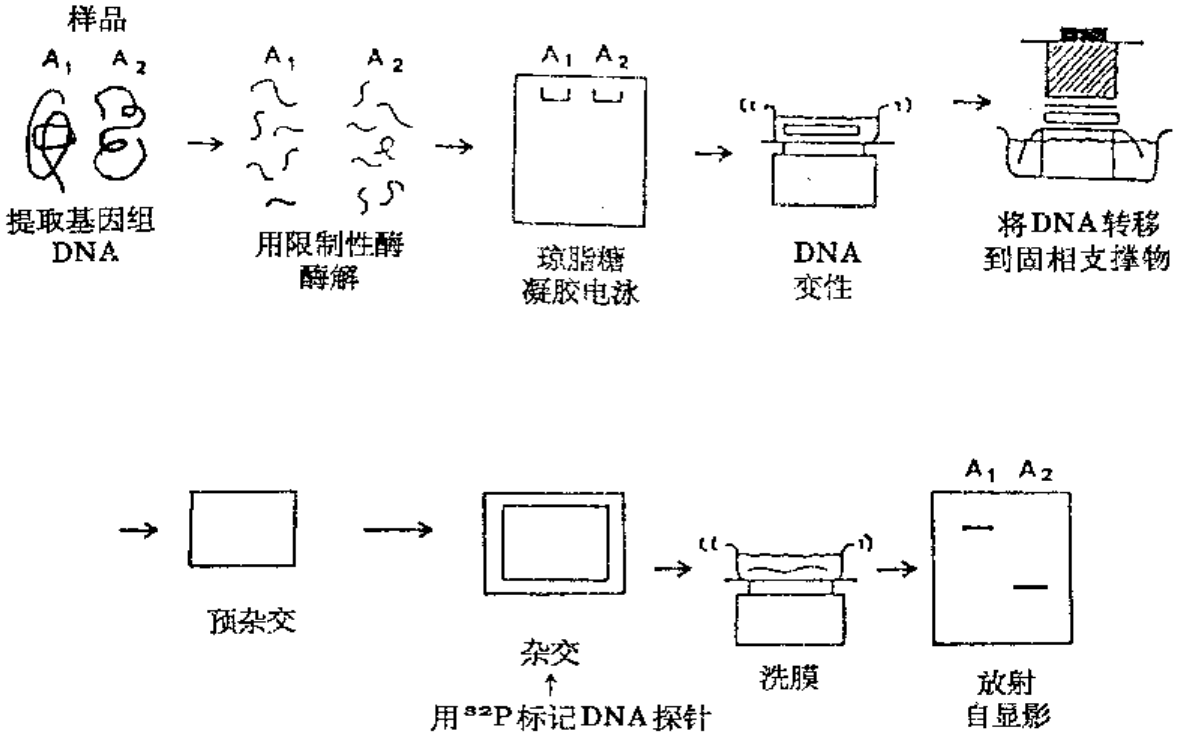


图 3.1 RFLP分析流程图

1. 原理

DNA多聚酶在催化DNA复制时需要一小段核苷酸序列作引物，在此基础上以解链后的单链DNA为模板，沿着5′到3′方向把核苷酸连续加在延伸中的游离3′羟基上以复制DNA。特定DNA序列需要特定的引物来诱导其复制的起始并决定复制的终止。由于不同物种或不同品种的基因组DNA序列差异很大，在复制特定的DNA序列时所需引物的核苷酸序列也不相同，同一引物可能使某一品种的DNA片段得以复制，而在另一品种中可能无法诱导。如此将引物所诱导复制的特定DNA片段采用PCR技术进行扩增，随后用电泳方法将扩增的DNA片段分离，即可使某一品种出现特定的DNA谱带，而在另一品种中可能无此谱带产生。因此，这种通过引物诱导及DNA扩增后得到的DNA片段多态性可作为一种分子标记。

2. 方法

(1) 引物合成

引物可通过DNA自动合成仪人工合成。专一配对的引物长度一般在20个核苷酸左右，核苷酸的序列应按照拟扩增的DNA模板的序列设定。

(2) DNA热循环扩增反应

扩增反应一般需经过数十次循环，每一循环分为以下几步：首先将DNA扩增混合液加热至94～96°C维持1～5分钟，使样品DNA解链；随后迅速将温度降低至36～55°C 1～2分钟，使引物与其同源的DNA片段退火结合；然后在1～2分钟内将反应温度升至70～72°C，

表 3.4　　RFLP标记的特点

遗传学	
	独特
	大量
	几乎涉及全基因组
	稳定遗传
	核基因组呈孟德尔遗传
	细胞器基因组呈母性遗传
	复等位
	共显性
	无上位性
	不受环境或发育模式的影响
	可检测杂交区域和相邻序列内部的变异
	可与高变区杂交
	可提供有关突变特性的信息
方法学	
	可在所有组织和所有年龄时期检测到
	DNA样本的搁置寿命较长
	探针/酶组合的数量无限
	可用杂合基因作探针
	利用单个探针可以检测多个基因座位，如多基因家簇、高变区
	任何探针都可以作为分离紧密连锁的标记的起点，通过染色体步行或跳跃，产生单倍型标记和括号类型标记，以简化分离
	通过染色体解析或分离可以获得特定染色体区域的探针
	通过插入可以形成额外的随机或位点特异多态性
限制	
	材料成本高
	费工精细
	天然存在的探针无法检测卫星DNA序列和高度重复的DNA序列

以激活混合液中的热稳定DNA多聚酶(Tag DNA多聚酶，此酶在96°C下活性稳定)；最后将温度保持在72°C持续0.5～2分钟，即可完成DNA片段的复制扩增(图3.2)。如此循环往复数十次即可使单一的特定DNA序列扩增10^5～10^7倍。扩增反应可在DNA热循环仪中进行。DNA扩增反应混合液主要由Tris-Cl (pH8.3～8.4)、KCl、$MgCl_2$、明胶、dATP、dCTP、dGTP、TTP、核苷酸引物、样品DNA以及热稳定DNA多聚酶(Tag DNA多聚酶)等组成。各成分的浓度则因供试物种的不同而有一定的差异。

(3) 电泳分离

扩增反应结束后，将DNA样品在乙醇中沉淀，然后重新溶解在TE缓冲液中。通过1.4～3%琼脂糖凝胶电泳分离并用溴乙锭染色10～15分钟，在紫外灯下即可直接观察到被扩增的DNA谱带，通过比较即发现不同样品DNA谱带的差异。

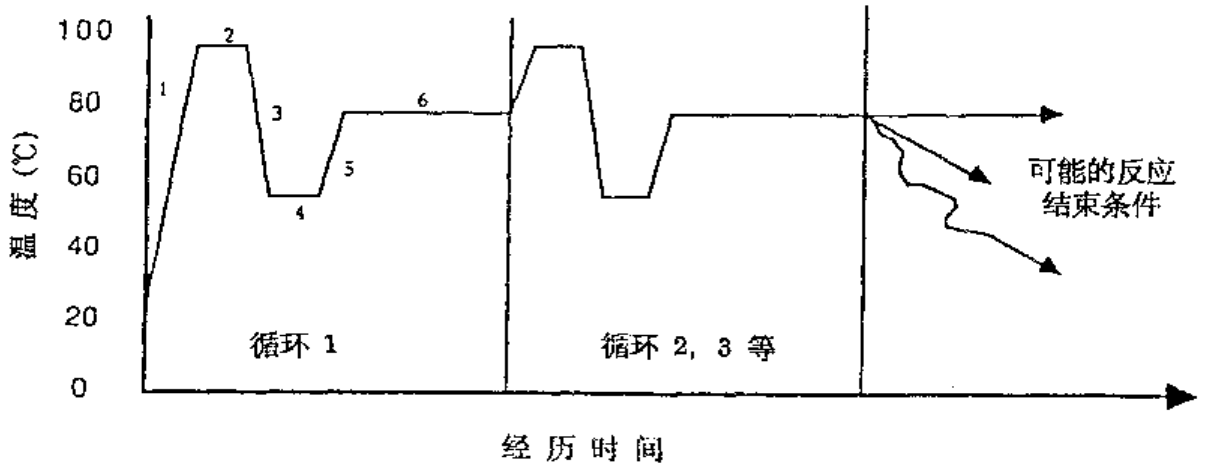


图 3.2　PCR扩增反应流程图

与RFLP 分析相比，用PCR 反应检测DNA 的多态性具有几个明显的优点。1. 分析时需要的基因组DNA 量少；2. 对DNA 制备的纯度要求不高；3. 程序简单，不必通过Southern 转移和分子杂交来显示片段长度的差异，因此分析的周期大大缩短。但是鉴于一般的PCR 反应的要求，在进行AFLP 分析前，必需根据基因组中被扩增的DNA 片段两端的序列设计或合成相应的引物，这是AFLP 分析的最大限制因素。不过，随着已知序列的基因数目的增加，AFLP 的适用范围将逐渐扩展。在实际应用中，考虑到AFLP 的上述优点，将RFLP 标记转换成AFLP 标记也是值得的。为此，应首先测定克隆的探针片段两端的部分序列，再根据序列设计和合成的相应的引物。由 RFLP转变而来的AFLP 有时候也称为基于 PCR的RFLP。在遗传分析时，AFLP 标记可能是共显性的，也可能是显性的。AFLP 分析有多种不同的形式，常用的引物都是专一性的，由此扩增的多态性称为专一扩增多态性 (Specific Amplified Polymorphism)，简称SAP。SAP 在概念上与 AFLP不完全等同。SAP既包括扩增片段长度的多态性，还包括扩增片段序列的多态性。不过这种在长度上没有差异但在序列组成上不同的片段不能通过一般凝胶电泳的方法予以区别，必须应用特殊的电泳分离技术。所谓的单链构象多态性 (Single Strand Conformation Polymorphism，简称SSCP) 就是在一种特殊的分离技术的基础上揭示这类多态性的。它的原理如下：生物大分子在单键旋转时由取代基团形成不同的立体结构而产生不同的构象，这种结构的改变不涉及共价键的破裂。DNA片段的双链成为单链后，在非变性的条件下会折叠成不同的空间构象，单链的这种构象会因为个别核苷酸的置换而改变，长度相同或相近的单链在电泳时由于构象不同而具有不同的迁移率，从而显示多态性。在进行SSCP分析时，应先将扩增后的DNA 片段变性成为单链，然后在非变性的聚丙烯酰胺凝胶中电泳。显然 SSCP扩大了 DNA多态性的检测范围。但由于在技术上比较复杂，在实际上很少用作常规的分子标记。另外，特殊的

电泳检测方法还有 DGGE (变性梯度凝胶电泳)和TGGE (温度梯度凝胶电泳), 它们也能揭示DNA序列组成上的差异, 不过这两种方法在分子标记作图中应用得更少, 在此不再赘述。另外, 还有一些基于非专一性引物的 AFLP 已发展成为专利技术 (见欧洲专利局专利号: 92402629.7)。

3.4.3 RAPD标记

在PCR 技术的基础上, Williams等(1990)采用随机核苷酸序列为引物扩增基因组DNA的随机片段, 获得了一种新的分子标记, 即随机扩增多态性DNA(Random Amplified Polymorphismic DNA), 简称RAPD。实际上RAPD是一种特殊形式的AFLP, 它与一般AFLP的最大区别在于用于扩增多态性DNA的引物不是专一的而是随机的。所谓的 RAPD 标记实际上是用随机序列组成的寡核苷酸通过专门的 PCR 反应扩增获得的长度不同的多态性DNA片段。

通常PCR 反应所用的引物都是特异序列, 扩增出的DNA片段也是特异的, 引物的长度一般在20个核苷酸左右。而RAPD 反应则是一种随机的扩增反应, 不仅引物的序列短而且核苷酸的组成和排列是随机的, 在PCR 反应中与基因组的配对位置取决于随机吻合的程度, 退火是在非严峻的条件(较低温度)下进行的。作为PCR技术的延伸, RAPD反应有其自身的特点, 区别于常规的PCR反应: 第一, 无需专门设计RAPD扩增反应的引物, 随机设计的、长度9～10bp的脱氧核苷酸序列均可适用。但是, 为了保证退火反应时双链的稳定性, G+C 含量应在40%以上。而常规的PCR反应, 必须通过已知的序列设计特定的引物。第二, 在每个RAPD 反应中, 可以只加入一个引物。通过一种引物在两条DNA互补链上的随机配对实现扩增。第三, 在最初的反应周期中, 退火温度较低, 一般为36℃。一方面保证短核苷酸链引物与模板的稳定配对, 另一方面允许适当的错误配对, 从而扩大引物在基因组DNA中配对的随机性, 提高对基因组DNA 进行分析的效率。

与RAPD 类似的随机扩增多态性的方法, 还有AP-PCR 和 DAF。AP-PCR是任意引物PCR (Arbitrary Primed PCR)的简称。在PCR 反应中使用的引物长度与一般 PCR反应中的引物相当, 但在反应开始阶段退火温度较低, 允许大量错配, 因此可引动具有随机性质的扩增。DAF是DNA扩增指纹图谱 (DNA Amplification Fingerprinting)的简称, 在PCR反应中, 使用的引物长度比 RAPD反应的更短, 为7～8个核苷酸。因此在扩增反应中有比RAPD更大的随机性, 扩增产生的 DNA带型也更多, 为此应使用聚丙烯酰胺凝胶电泳, 并用银染检测 DNA 片段。由RAPD、AP-PCR 和 DAF一起组成的多态性分析方法统称为 MAAP。MAAP是任意扩增多带谱 (Multiple Arbitrary Amplicon Profiling) 的简称, 不过在遗传作图中用得最多的还是 RAPD标记。

RAPD 标记可以转化为 RFLP 标记。将 RAPD 片段从凝胶上分离后, 克隆在大肠杆菌的质粒载体上, 即可进行 RFLP 分析, 如不是重复序列, 就成为 RFLP 标记了。与RFLP相比, RAPD标记有如下三个特点(表3.5, Waugh和Power 1992): (1) 合成一套引物, 可以用于不同生物基因组的分析, 而相对来说, RFLP标记具有一定的种族特异性, 不能广泛使用。 (2) 大大减少了多态性分析的预备性工作, 如制备克隆、同位素标记、Southern

表 3.5　RFLP 和 RAPD 技术的特性比较

	RFLP	RAPD
分布	普遍存在	普遍存在
遗传	简单孟德尔 稳定 共显性	简单孟德尔 稳定 显性/零等位
多态性程度	高	高
等位基因变异的检测	可以	不能
一次可检测的基因座位数	1～3	1～10
被检测的基因组部位	一般为低拷贝、编码区	整个基因组
所有组织均可检测	是	是
技术难度	中等	低
可靠性	高	中等/高（偶有假像）
所需DNA的质量	相对较纯	粗制
所需DNA的数量	2～10 μg	10～50ng
同位素	需要	不需要
所需探针类型	物种特异性、低拷贝的基因组DNA或cDNA克隆	9或10个随机寡核苷酸
成本	高	高

吸印、分子杂交等，而且使用的DNA量极少。(3) 绝大部分的RAPD标记不能区别杂合型和纯合型基因组的RAPD反应产物，因此是显性标记。虽然RAPD标记有上述特点，但也有其不足的一面，即每个标记提供的信息量少，RAPD的检测受反应条件的影响较大，重复性差等。因此，在构建 RFLP图谱时，需将 RAPD 标记克隆后，转变成 RFLP 标记来使用。由于RAPD分析比RFLP 分析简便快捷，RAPD标记已开始在植物基因定位和作图研究中广泛地应用。

目前， RAPD 标记已广泛地应用于遗传作图，基因的快速定位，特殊染色体区段的鉴定和分离，以及动植物育种等许多方面。Martin等 (1991) 以番茄的近等基因系为材料，

表 3.6　　c-RFLP, r-RFLP 和 RAPD技术的实验步骤

c-RFLP	r-RFLP	RAPD
冻干、碾磨、DNA的分离 冻　干 碾　磨 氯仿抽提 酚抽提 乙醇沉淀 DNA 洗涤 DNA 含量分析		
限制性酶切 酶　切 沉　淀 再悬浮、加样		PCR 扩增
琼脂糖凝胶电泳 制胶、电泳 照　像		
Southern转移		
探针标记 PCR 扩增	寡核苷酸标记	检　测 琼脂糖凝胶电泳 照　像
杂　交		
洗膜和检测 洗　膜 检　测 曝　光	洗　膜 洗　膜 曝　光	
除 探 针		

注　　c-RFLP: 化学发光为基础的RFLP; r-RFLP:以同位素标记为基础的RFLP

得到了一个与细菌病抗性基因（ *pto*）连锁的 RAPD 标记，并定位于番茄的第5条染色体上。Michelmore等(1991)利用莴苣的分离群体，获得了一个抗霉菌病的RAPD标记。这些研究成果，显示了RAPD技术极大的实用价值。

有关c-RFLP, r-RFLP 和 RAPD技术实验步骤的相互比较见表3.6 (Ragot和Hoisington 1993)。

3.4.4 重复序列

在真核生物基因组中，除了基因的编码序列和调控序列外，还有许多未知功能的重复序列。植物基因组的重复序列一般在50%以上，在某些植物如小麦基因组的重复序列多达80%以上。重复序列按其在染色体上的分布方式，可分为散布重复序列(Dispersed Repeats Sequence)和串联重复序列(Tandem Repeats Sequence, TRS)。散布重复序列的拷贝数很多，而且在重复单位之间彼此常有序列的变化，在 RFLP分析中不能显示出清晰的带型（通常为模糊一片），因此难以用作 RFLP标记。在人类基因组中有一个Alu重复序列的家族，重复单位约300bp左右，它们广泛地散布在人的基因组中，而且有丰富的多态性。目前已可以用专门设计的 PCR方法揭示这种Alu多态性，称为Alu-PCR。通过Alu-PCR可提供这类重复序列的分子标记。不过，迄今在植物中尚未见利用散布重复序列提供分子标记的报道。

串联重复序列按重复单位的大小，可分为卫星序列(Satellite)、小卫星序列 (Mini-satellite)和微卫星序列(Microsatellite)三种。其中卫星序列的重复单位很大，一般分布在染色体的异染色质区，也难以采用分子杂交或PCR的方法揭示其多态性。然而后两种序列因为重复单位较小，由重复单位的序列差异和重复单位的数目变化，可形成非常丰富的多态性。目前已发展了多种方法，利用小卫星的多态性进行基因组的指纹分析，开拓微卫星的多态性作为遗传作图的分子标记。事实上，在人类和哺乳动物的分子连锁图谱中，微卫星已成为取代RFLP的第二代分子标记而被广泛使用。

1. 小卫星DNA：是一些重复单位在11～60bp，总长度由数百到数千碱基组成的串联重复序列，主要存在于染色体的近端粒 (Telomere)处，在不同个体间存在串联数目的差异，因而可进行多态性分析和染色体定位 (Nakamura等 1987, 1988)。由小卫星DNA 组成的染色体座位具有丰富的多态性，这是由重复单位的数目变异造成的，这种多态性亦可称为VNTR (Variable Number of Tandem Repeat Locus, VNTR, 见Nakamura等1987)。VNTR 可能成为一种广泛应用的复等位标记。对小卫星DNA的多态性鉴别可用重复单位的同源序列作为特异探针进行 RFLP 分析，因为在每个染色体座位上的串联重复序列中有重复出现的限制性内切酶靶序位。在酶解后，会出现一条主带，它们可以用分子杂交和自显影的方法加以分辨。检测用的探针可在基因组文库中筛选(Nakamura等 1987; 1988)，也可人工合成。

2. 微卫星DNA：微卫星DNA 又称为短串联重复 (Short Tandem Repeat)，简称STR，它们的多态性又称为简单序列长度多态性 (Simple Sequence Length Polymorphism)，简称SSLP。这是一类由几个核苷酸（一般为1～5个）为重复单位组成的长达几十个核苷酸的串联重复序列。同一类的微卫星DNA可分布于整个基因组的不同位置上。例如，在人类基因组中广泛存在的一种微卫星DNA是 $(dC-dA)_n(dG-dT)_n$，由于重复的次数不同以及重复

程度的不完全而造成了每个座位的多态性(Weber 1990b)，这种多态性的信息量是比较丰富的。由于在每个微卫星DNA 两端的序列多是相对保守的单拷贝序列，因而可根据两端的序列设计一对特异的引物，扩增每个位点的微卫星序列，再经聚丙烯酰胺凝胶电泳，比较扩增产物的长短变化，即可显示不同基因型的个体在每个微卫星DNA 位点(Site) 的多态性；一般微卫星标记在遗传上是共显性的。由于微卫星DNA 蕴含的多态性很大，目前已成为人类遗传连锁图谱研究的热点。最近已建成一张完全由微卫星DNA 标记组成的人的第20对染色体的连锁图，标记数达26个，远远超过已定位的RFLP 标记数。但微卫星DNA作为一种标记，其方法学上的缺点是必须针对每个染色体座位的微卫星，发现其两端的单拷贝序列以设计引物，这给微卫星标记的开发带来了一定的困难。但一旦开发出一套某种生物的微卫星标记，其他的研究工作者就受益无穷。

3.4.5 STS

STS是序列标记位点(Sequence-Tagged Sites)的简称。它是由一段长度为200～500bp 的序列所界定的位点，在基因组中只出现一次。前述任何单拷贝的多态性标记都可以作为基因组的界标(Landmark) 转变为STS 标记。转变的前提是测定长度适合的单拷贝片段的两端序列，设计一对专一扩增的引物(长度为20个核苷酸左右)，用PCR 方法显示STS 的特异片段。最富信息和多态性的STS标记可能是扩增含有微卫星重复顺序的DNA区域所得到的STS标记(Tautz 1989; Weber和May 1989)。STS类标记的优点在于其共显性遗传方式，因而很容易在不同组合的遗传图谱间进行标记的转移。Olson等(1989)倡导利用STS 标记使人类基因组物理作图标准化。在遗传图谱上定位的STS 标记是沟通基因组遗传图谱和物理图谱的中介。如同微卫星标记一样，STS 的开发也依赖于测序，但一旦开拓，同行受益。

上述反映多态性的所有分子标记，都是基于基因组DNA水平的差异和相应的检测技术，只有两者的结合才能提供尽可能多的遗传标记。DNA的变异是客观的，自然的，也是丰富的，而技术的进步则是人为的。因此，随着新技术的发展，可以开发的分子标记将越来越多。而且，随着分子标记的拓展和广泛应用，遗传图谱的使用价值将日益提高，最终将成为借助于生物技术利用和改造生物的依据。与此同时，遗传图谱精密度的提高，也必然会促进各种生物基因组研究向深入发展。

3.5 各类遗传标记的特点与利用

同工酶和RFLP 等分子多态性的发现，已使遗传作图的研究上了一个新台阶而进入微观的水平。不过，尽管我们可以根据大量同工酶或RFLP 之间的连锁关系确定各自的连锁群，但要将这些连锁群与特定的染色体联系在一起，最简单的方式还是要求助于经典的遗传标记。如许多同工酶基因的定位最初就是依靠细胞学标记进行的。在水稻中将RFLP 连锁群定位于特定的染色体就是采用经典的三体为标记进行的，而一些同工酶基因的定位则采用了经典的形态标记（如糯性基因）。在大麦、玉米、番茄中也都成功地利用三体进行了同工酶基因的定位。而且在有些物种中，利用端着丝点染色体可以确定同工酶基因的染

色体臂位。这些已定位的同工酶基因就成为其它基因定位的遗传标记。事实证明，较早建立RFLP 连锁群的物种大多数是经典遗传研究和基因定位比较完善的物种。

大多数形态标记具有典型的形态特征，比较容易识别和观测，因而便于确定它们与其它性状的关系。同时，由于试验设备简单，分析成本低廉，使它们不仅成为大多数物种连锁群建立的开创性标记，而且在其它分子标记的定位中也起着重要作用。但其缺点是有些形态标记与其它有害或不利性状相连锁或具有不良的多效性。如水稻的大多数矮秆标记基因使营养生长削弱，伴随着多蘖、短穗和低产等多效性，使分析和利用受到限制。此外，利用形态标记进行数量基因定位还需要严格控制试验环境，以使不同个体间形态特征的差异更为明显。细胞学标记克服了形态标记易受环境影响的缺点，但这种标记材料的产生需要花费大量的人力进行培育和选择，一些物种忍受染色体结构和数目变异的能力较差而难以获得相应的标记材料；某些物种虽有标记材料，但常常伴随着标记对生物有害的表型效应，而且有时观测和鉴定比较困难。与经典的形态标记和细胞学标记相比，从微生物到高等真核生物DNA 分子多态性是普遍存在的，它是自发产生的，在同一基因座位上有较多的等位基因，因而在应用范围上所受限制较少，这对于形态标记和细胞学标记发展不够充分的物种尤为重要；分子多态性对生物本身通常无害，对重要经济性状很少有次级效应，因而具有很大的选择余地；分子多态性通常能在个体、组织、器官和细胞水平测定，且不受环境影响，遗传分析可以在生物发育的各个时期进行，这给研究工作提供了极大的方便；分子多态性基因座位的等位基因为共显性，在分离群体中能区别所有可能的基因型，或至少是显性(而不是隐性)的，因而可以在所有杂交组合中使用。分子标记所具有的这些优点给基因定位尤其是数量性状的基因定位展示了美好的前景。但分子标记特别是 RFLP 分析过程复杂，分析和检测需要特定的仪器设备，单个样本的分析成本较高。

不同遗传标记所具备的特点决定了它们在基因组研究中具有物种、试验材料和分析时期的特异性。从已有的研究基础考虑，对于那些连锁群尚未建立的物种，要通过对各种遗传多态性进行筛选，寻找具有丰富多态性的标记着手建立连锁群。当具有丰富的形态多样性可以利用时，形态标记就成为一种方便而有效的方法。在条件许可的情况下，通过检测分子标记来建立连锁群是一种快速而有效的方法。对于连锁群已基本饱和的物种，可着手利用分子标记进行分子水平的基因定位和基因精细结构的分析，使基因组研究达到一个新水平。

从生物的繁殖方式考虑，异交生物的纯合化比较困难，采用分子标记进行遗传研究可以减少产生近交系的麻烦。从染色体倍性水平上考虑，多倍体生物对染色体畸变有较强的耐性，可以利用非整倍体及结构变异体等细胞学标记。某些同工酶只在植株发育的特定时期表达，因而这些酶的遗传研究就受其表达时间的限制；而从细胞的全能性出发，一个细胞的DNA可以代表它所来自的一个母体，因此以RFLP等遗传标记的研究可以在生物发育的任何时期进行，而且可以将DNA样本贮存起来在方便的时候进行分析。当具有广泛的近缘物种可以利用时，采用被研究物种与其近缘野生种之间的杂种可能包括大量有差异的等位基因，使单个杂交组合所能分析的基因座位数大大增加，从而提高分析效率。

随着各种遗传标记的逐步完善和发展，现在已进入综合应用各种遗传标记进行基因组研究的新阶段。综合应用可以扬长避短，提高基因组研究的整体效率。例如，一些过去只

靠一种标记无法完成的基因定位，现在通过几种遗传标记的综合应用而得以实现。几种标记的同时应用，其结果可以相互验证，使基因定位更加准确。在大豆中，采用3个形态标记和4个同工酶标记同时进行连锁分析，发现了过去采用单一标记未能定位的叶绿素缺乏突变体（Y9）基因的连锁关系，并建立了一个新的连锁群。在水稻中，由于形态标记和复等位基因材料的贫乏，曾经过多年努力定位了三个同工酶基因，而水稻初级三体的建立为同工酶基因定位提供了有用的工具，在利用三体为标记的一次试验中就确定了10个同工酶基因座位所属的染色体。

遗传标记的发展是随着人类对于基因由现象到本质的认识而由形态标记向分子标记逐步发展的过程。在这一过程中，传统的形态标记和细胞学标记是遗传标记发展的基础，而生化标记和分子标记则是遗传学和生物化学的发展导致遗传标记发展的必然结果。现代分子标记的出现促使遗传图谱构建向高效率、高精度以及揭示基因精细结构的方向发展，从而为遗传学注入了新的活力。

第 4 章　遗传图谱构建

利用遗传标记构建连锁图谱的一般步骤包括：选择用于作图的适合的遗传标记；根据遗传材料之间的多态性，确定用于产生作图群体的亲本和组合；培育具有大量遗传标记处于分离状态的分离群体或衍生系；作图群体中不同个体或品系标记基因型的确定；标记之间连锁群的构建。本章只对其中一些问题作详细的论述。

4.1　作图群体的建立

连锁图谱构建的理论基础是染色体的交换与重组。在细胞减数分裂时，非同源染色体上的基因相互独立、自由组合，同源染色体上的连锁基因产生交换与重组，交换的频率随基因间距离的增加而增大，因此可用重组率来揭示基因间的遗传图距，图距单位用厘摩(centi-Morgan, cM)表示，1个cM的大小大致符合1%的重组率。

连锁图谱构建的遗传材料是分离群体，用于作图的分离群体又称为作图群体。为作图群体选择合适的亲本和适当的杂交组合类型，是成功和高效作图的关键。

4.1.1　亲本的选择

亲本的选择直接影响到构建连锁图谱的难易程度及所建图谱的适用范围。既然DNA的序列差异是分子图谱构建的基础，就应该选择亲缘关系远而多态性大的品种或材料作为分离群体的亲本。然而亲本间的差异过大，杂种染色体之间的配对和重组会受到抑制，导致连锁座位间的重组率偏低，降低了所建图谱的可信度和利用的范围；严重的会降低杂种后代的结实率，甚至导致不育，影响分离群体的构建。因此，在选择亲本时，既要注意亲本间的差异性，也要考虑其后代的可育性。然而，由于各种原因，仅用一对亲本的分离群体建立的遗传图谱往往不能完全满足基因组研究和各种育种目标的要求，解决的办法是选用几个不同的亲本组合，分别进行连锁作图，以达到相互弥补的目的。除去双亲的多态性水平外，配制组合前还应对亲本及其F_1杂种进行细胞学鉴定。若双亲间存在相互易位，或多倍体材料（如小麦）存在单体或部分染色体缺失等问题，那末其后代就不宜用来构建连锁图谱。

亲本之间的亲缘关系究竟以多远为宜？这因物种而异，一般而言异交作物的多态性高，自交作物的多态性低。例如，玉米的多态性极好，一般品种间配制的群体就可成为理想的RFLP作图群体，而番茄的多态性较差，因而只能选用不同种间的后代构建RFLP连锁图，水稻的多态性居中，美国康乃尔大学S. D. Tanksley实验室1988年发表的RFLP连锁图谱是以籼稻和爪哇稻之间的杂交组合为基础构建的(McCouch等 1988)。

4.1.2　组合的配制

组合配制过程中最关键的一条就是应尽量保证所用群体起源于同一F_1个体。大多数作物的繁殖系数都较高，这一点并不难做到。如果非用若干个F_1不可时，应保证将不同F_1及其所衍生的后代分开保存。这是因为，无论亲本自身纯度如何，每个F_1都是一个雄配子与一个雌配子结合的产物。所以由同一F_1个体衍生的后代在每一基因座位上只可能有两个等位基因发生分离。相反，如果所用分离群体不是起源于同一F_1个体，该群体的某些基因座位就可能存在两个以上等位基因的同时分离，因而给结果的分析带来困难。这一点对异花授粉作物尤其重要。

另外，如果可能，应尽量选用纯度比较高的材料作亲本。同时对用作亲本的植物应尽可能多地采集样本，并自交留种。

4.1 3 作图群体的类型

这里简单地介绍常用于作图的4类作图群体，详细的讨论见第5章。

构建RFLP连锁图谱可以选用不同类型的分离群体。不同类型群体各有其优缺点，因此应结合具体情况选用。

1. 单交组合所产生的F_2代或由其衍生的F_3、F_4家系

迄今为止，大多数 RFLP连锁图谱是基于单交产生的F_2群体。除自交不亲和的材料外，这种群体易于配制，且不需很长时间。但其致命弱点是由F_2单株提供的材料（种子或叶片）有限，很难用其进行连续性研究，也无法满足不同实验室对同一材料合作研究的需要。现行的一种补救办法是利用F_2代单株衍生的后代家系。这时就需要选取同一家系中的若干个体（或其混合物），才能够比较准确地推断出F_2个体的基因型。这种做法不仅加大了工作量，而且会出现取样误差。

2. 重组近交系 (Recombinant Inbred Lines, RIL)

由于这种群体是由连续多代自交或姊妹交产生的，这使得染色体间重组机会增加，因而可以用来更精确地定位那些连锁紧密的位点 (Burr和Burr 1991)。这种群体的另一个优点是可以连续不断地提供大量的种子或叶片等材料，因而可以满足不同时间、不同地点、不同实验室研究的需要。其缺点是配制这种组合所需时间较长。

3. 加倍单倍体 (Doubled Haploids, DH)

与重组近交系一样，加倍单倍体同样可以大量地提供不同的试验材料，而且这种群体的产生又不需经过许多世代。只需将花药培养或其它途径得到的单倍体材料加倍就可以得到稳定遗传的材料，每个材料都可象一般品种一样地进行繁殖。但不同基因型对花药培养的反应可能是不同的，有人因此怀疑用这种经过培养选择的群体所构建的连锁图谱的可信度。对此，我们曾对自己试验中所得的一组数据进行两种不同的处理：删除其中一个基因座位一部分纯合基因型，或删除其中一部分杂合基因型。经过这种不同处理后所得连锁图谱与未经处理所得连锁图谱的基因座位顺序是完全一致的，基因座位间的距离也仅有微小差异，而这些差异的一部分还会是由于删除数据后群体变小所致。就是说，相对于不同组合之间所得连锁图的差异而言，由于这种选择所造成的差异可能是有限的。

4. 回交和三交(复交)群体

对于自交不亲和材料，或在研究某种特殊问题（如雌雄配子交换率的差别等），就需

要利用回交或三交(复交)后代。与单交组合所产生的F_2代群体一样，回交和三交所能提供的材料也是有限的。另外，除自交不亲和材料外，一般回交群体都或多或少地含有自交的成分。但是如果采用三交，则可容易地区别后代材料的真伪。相对于三交，利用回交后代也有其优点，那就是可将回交的正反交做成同质的［例如 (A×B) ×A与A× (A×B) ］，这样更有利于比较核基因的差别。而利用三交则无法做到这一点。

上述作图群体可按其遗传稳定性分为两大类：

1) 非固定性分离群体(F_2、F_3、F_4 BC群体、三交群体)

2) 固定性或永久性分离群体(由基因型不同的同质纯合品系组成)。

上述两类群体是植物中常用的作图群体，它们之间的关系可以图4.1所示。

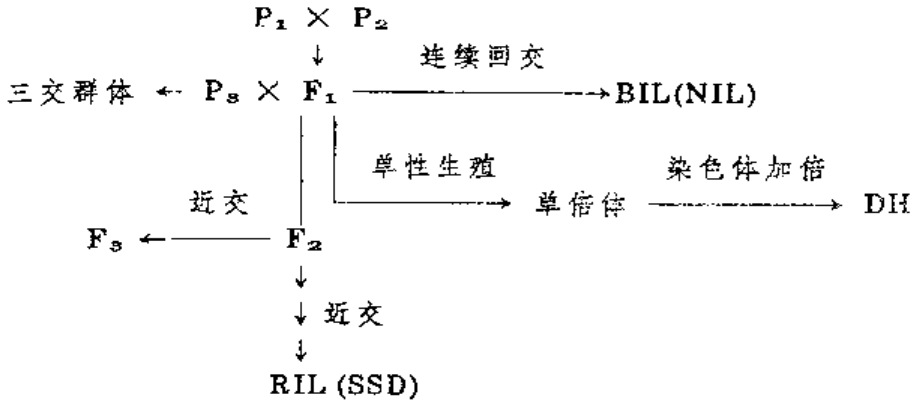


图 4.1　几种作图群体之间的关系

水稻中已有的作图群体包括：

F_2：　IR34583-19-3-3(籼稻)×Bulu Dalam(爪哇稻) (McCouch等1988)
Kasalath (籼稻)×FL134 (粳稻)　(Saito　等　1991)
Zhai-Ye-Qing 8 (籼稻)/ Jing-Xi 17 (粳稻) (徐云碧等 1994b)

BC:　BS125 (籼稻)×(BS125×WLD2 [野生稻 *O. longstaminata*])
(Tanksley等 1992)

DH:　IRAT177(粳稻)×Apura(籼稻)
Zhai-Ye-Qing 8 (籼稻)/ Jing-Xi 17 (粳稻) (Zhu等1993, Xu等1994)
Gui630 (籼稻) × 02428 (粳稻) (中国科学院遗传所)

SSD:　IRAT177×Apura
CO39(籼稻)×Moroberekan (粳稻) (Wang等1994)

RIL:　Lemont×特青(籼)

4.2　连锁作图的理论基础

4.2.1　Sutton-Boveri染色体遗传理论

遗传因子位于染色体上的理论是1903年分别由W. S. Sutton和T. Boveri提出的，他

们将染色体看作是孟德尔基因的物理载体。其基本要点如下：

(1) 体细胞核内的染色体成对存在，其中一条来自雌亲，一条来自雄亲，成对染色体的两个成员是同源的。

(2) 每条染色体在个体的生命周期中均能保持其结构上的恒定性和遗传上的连续性，因而在个体的发育过程中起着一定的作用。

(3) 在减数分裂中，同源染色体的两个成员相互配对，随后又发生分离，而走向细胞的两极，从而形成两个单倍体性细胞。

4.2.2 基因重组和连锁

现在我们知道，位于同一染色体上的基因在遗传过程中一般倾向于维系在一起，而表现为基因连锁。连锁遗传的基因是顺序排列在一对同源染色体上的一些非等位基因。它们之间的重组是通过一对同源染色体的两个非姊妹染色单体之间的交换来实现的。

设某一对同源染色体上存在 $A-a$，$B-b$ 两对连锁基因，现有两个亲本 P_1 和 P_2，它们的基因型分别为 $AABB$ 和 $aabb$，两亲杂交产生 $AaBb$ 双杂合体(图4.2)。

F_1 在减数分裂过程中应产生4种类型的配子，其中两种为亲型配子 AB 和 ab，两种为重组型配子 Ab 和 aB (图4.2)。由于 $A-a$ 和 $B-b$ 位于同一染色体上，要产生重组型配子必须在这两个基因的连锁区段上发生交换。重组型配子所占的比例取决于减数分裂细胞中发生交换的频率。即使所有减数分裂的细胞中，在这两对基因的连锁区段上都发生交换，则所产生的重组型配子也只有50%。

重组型配子占总配子的比例称为重组率。用 r 表示。则非重组率为1- r。 重组率的

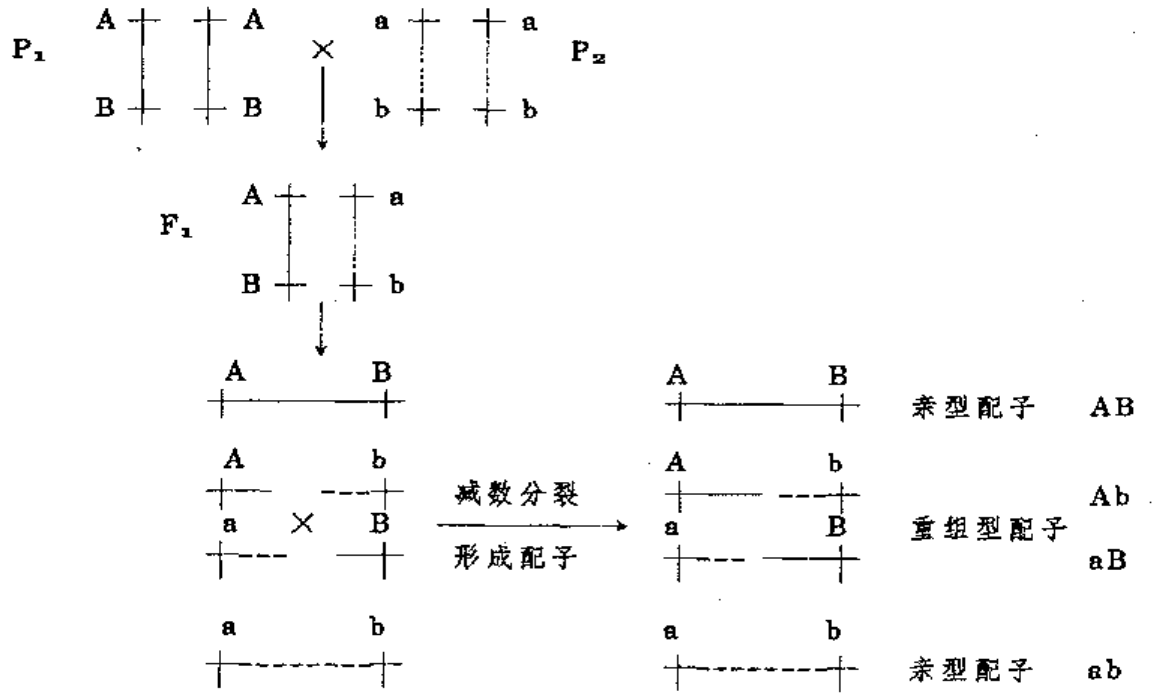


图 4.2 由杂交组合 $AABB\times$ aabb 所产生的亲型配子和重组型配子

高低取决于交换的频率，而两对基因之间的交换频率取决于它们之间的直线距离。重组率从完全连锁时的0%到完全独立时的50%。

当两对基因完全连锁时，它们就表现为一因多效，其表现类似于一对基因，而两对基因完全独立时，它们之间可以自由重组，F_1 四种配子的比例为1:1:1:1。在上述双杂合体所产生的4种配子中，其比例如表4.1。

表 4.1　基因连锁和独立状态下亲本组合 *AABB*×*aabb* 所产生的配子类型及其比例

配　子	*AB*	*Ab*	*aB*	*ab*
连锁	$(1-r)/2$	$r/2$	$r/2$	$(1-r)/2$
完全连锁($r=0$)	1/2	0	0	1/2
完全独立($r=0.5$)	0.25	0.25	0.25	0.25

产生两对基因的杂合体 *AaBb*，有两种方式。一种是上述两个基因座位都是显性基因聚集于一个亲本中，而隐性基因聚集于另一亲本中。这种组合称为相引组合，另一种是一个座位的显性基因与另一个座位的隐性基因分布于一个亲本中，而另一个亲本与之相反，即 *AAbb*×*aaBB*。称为相斥组合。对于分子标记，其基因座位上的等位基因通常没有显隐性之分而表现为共显性。这时两个等位基因只有书写形式上的不同。

上述重组率是以非姊妹染色单体之间在 *A*－*a* 和 *B*－*b* 连锁区段之内发生的一次交换(单交换)为基础的。但是，非姊妹染色单体之间的交换并不都是如此简单。当 *A*－*a* 和 *B*－*b* 之间相距较远时，两条非姊妹染色单体之间可能在其连锁区段发生两次交换(双交换)。这种双交换把单交换所产生的重组型配子回复到了亲型配子。但是，两对连锁基因之间的双交换实际上都不会降低到以单交换为基础所估算的重组率。原因是两对连锁基因之间双交换的形式是多种多样的。归纳起来有四种，即相互双交换，相补双交换和两种形式的相对双交换。发生相互双交换的性母细胞虽然只产生亲型配子，但发生相补双交换的性母细胞只产生重组型配子。两者结合在一起使得亲型配子和重组型配子各占50%的比率未变。至于发生两种相对双交换的性母细胞，它们都是各自产生50%的亲型配子和50%的重组型配子，同发生一次单交换的性母细胞所产生的配子类型比例没有区别。因此，即使是某些性母细胞内两对连锁基因之间发生了各种形式的双交换，只要其机会均等，就不会改变以单交换为代表的亲型和重组型配子各占50%的比率。

4.2.3　影响交换和重组的因素

人们很早就认识到交换和重组是受许多因素影响的。这些因素主要包括：

1. 环境因素

许多环境因素影响交换和重组，其中包括温度、水分、营养、射线、化学药品等。早在1917年，Ploug就报道了果蝇第二染色体上的钝刚毛基因（*B1*）与残翅基因（*vg*）之间的重组率随培养温度的增加而增加。在25℃时为6.1～6.5%，在36～38℃时为8.1～

9.6%。Lawrence (1965)以百合和鸭跖草为材料，用射线在S期进行处理，使重组率低于对照(未处理者)，而偶线期至粗线期处理则重组率高于对照。

2. 性别

果蝇的交换只发生在雌蝇而不发生在雄蝇，家蚕则刚好相;豌豆的两性交换值相等，鼠类雌体的交换值比雄体大；而鸽的雄体交换值比雌体大。实际上，凡是性别为性染色体所控制的生物，异配性别的个体中一般总是较少发生交换。霍尔丹定律就是反映生物的这一特点的。该定律指出，凡是较少发生交换的个体一定是异配性别个体。

玉米是雌雄同株异花的植物，雄花序和雌花序虽无染色体上的异型，但某些连锁基因的重组率却是雄花序高于雌花序。Rhoades(1944)研究了玉米第五染色体上基因重组率与雌雄花序的关系，结果如表4.2。

表 4.2　玉米雌雄花序基因重组率的差异

	a_2-bm	a_2-bt	bm-pr	bt-pr
雄花序	15.6	16.5	33.8	35.4
雌花序	9.1	9.7	22.4	30.3

a_2 —无花青素基因；bm—褐色中脉；bt—胞化胚乳；pr—红色糊粉层

3. 着丝粒和B染色体

着丝粒附近的交换频率受环境影响最大，如温度、年龄等。由于着丝粒附近是异染色质部分，是高度重复的随体DNA，一般交换率降低。

B染色体(超数染色体的一种)存在于大多数动植物细胞中，它比A染色体(正常染色体)为小，而且和任何A染色体都不是同源的，因此在减数分裂过程中B染色体一般不和A染色体联会。但是A染色体的交叉频率却由于B染色体的存在表现增加或减少。Rees(1967, 1968)研究指出，在玉米中，随着B染色体数的增加，玉米性母细胞减数分裂第一前期中的交叉频率相应增加，而对黑麦草属，B染色体的作用则相反，随B染色体数的增加，交叉频率相应降低。

4. 染色体异常

许多研究结果证明，在倒位杂合体中，倒位可有效地抑制倒位区域内的交换。染色体相互易位同样可以起到保护某种基因组合而减少重组的作用。一个极端的例子是拉马克月见草（*Oenothera lamarckiana*）的一些易位染合体，涉及到7对染色体，所有的染色体成为一个“连锁群”，很少能通过交换而在这个“连锁群”间发生基因重组。

5. 重组的基因控制

基因型改变为突变型后可能对染色体交换产生影响。如在果蝇、人、小麦、玉米、真菌中链孢霉、曲霉、酵母菌等，以及在λ噬菌体中，都存在抑制交换或促进交换的突变型。交换频率的基因控制包括单基因控制和多基因控制。在单基因控制方面，动物(果蝇和人)和高等植物(小麦和玉米)中都有报道。在多基因控制方面，Rees (1955)发现一个自

然传粉的黑麦品种经过5～6代连续自交之后，不同自交系的交叉频率逐渐下降到不同的稳定水平，将两个交叉率已下降不同水平的稳定自交系进行杂交，该杂种的交叉频率大幅度提高，显著地超过了两个亲本自交系。再将该杂种连续自交，则F_2、F_3的交叉频率又逐代下降，其平均值恰恰近似于两个亲本交叉频率的平均值。很明显，加性效应的多基因对交叉频率的遗传起决定性作用。

4.2.4 遗传图距的相对性

根据染色体交换所确定的遗传图距除了受上述因素的影响外，还存在一些具有普遍意义的因素，导致遗传图距的相对性。首先，对于不同生物，每一cM图距所代表的实际物理距离（碱基对数）存在很大差异。例如在噬菌体T_4中每一cM图距相当于200bp，而在酵母中相当于5,000bp，在果蝇中则相当于500,000bp。总的趋势是生物越简单，每一cM图距所相当的碱基对数就越少（表4.3）。其次，同一种生物特定基因间的遗传图距因基因型不同而改变，甚至由同一对亲本所产生的不同群体间也存在很大差异。第三，同一条染色体的不同区域上，由于同样长度的染色体片段发生交换的频率存在差异而导致遗传图距的差异。例如着丝粒附近染色体的交换受抑制，因而所估计的遗传图距将小于其实际距离。

表 4.3 不同生物中单位图距所相当的碱基对数

物　种	基因组大小(bp)	遗传图距(cM)	bp/cM
噬菌体T_4	1.6×10^5	800	200
大肠杆菌	4.2×10^6	1,750	2,400
酵　母	2.0×10^7	4,200	5,000
真　菌	2.7×10^7	1,000	27,000
线　虫	8.0×10^7	320	250,000
果　蝇	1.4×10^8	280	500,000
小　鼠	3.0×10^9	1,700	1,800,000
人　类	3.3×10^9	3,300	1,000,000

4.3 遗传图谱的制作

4.3.1 分子标记分离数据的收集与处理

遗传连锁分析的第一步是获得不同个体的DNA多态性资料。由于各种分子标记最后显示的形式，都是电泳分离的带谱，因此，从群体中收集分子标记分离数据的关键是将电泳的带型数值化。我们以RFLP为例，说明分子数据的收集和处理。

各个体的DNA经限制性酶切之后进行凝胶电泳，使不同长度的片段分开。将这些片段经Southern印迹转移到尼龙膜上，然后与放射性同位素标记的探针杂交，经放射自显影就得到不同带型的各个体数据（表4.4；表4.5）。

在表4.4中，两个亲本各显示一条带，而F_2群体中不同个体的带型有三种，即P_1型、

P_2型和杂合体类型。如果我们将含有P_1带型的F_2个体赋以值1，P_2带型的个体赋以值3，杂合体赋以值2，于是我们可以得到数据化的F_2 RFLP资料如表4.5。

表 4.4 亲本P_1、P_2及F_2个体的DNA经RFLP分析所得放射自显影结果示意表

	P_1	P_2	F_2 1	2	3	4	5	6	7	8	9	10	11	12	...
RG303		-	-	-	-	-	-	-	-		-		-	-	...
	-		-		-		-		-	-	-	-		-	...
RG196		-	-	-	-	-	-	-	-		-	-	-	-	...
	-		-		-		-		-	-	-	-		-	...
...	...														...

RG为随机基因组克隆的探针编号

表 4.5 P_1、P_2和F_2群体各个体RFLP分析结果的数据化

	P_1	P_2	F_2 1	2	3	4	5	6	7	8	9	10	11	12	...
RG303	1	3	2	3	2	3	2	3	2	1	2	1	3	2	...
RG196	1	3	2	3	2	3	2	3	2	1	2	2	3	2	...
...	...														...

假设全部试验共有120个F_2单株，利用100个探针进行分子杂交。这样我们就可以获得一个由100行、120列数据组成的RFLP数据矩阵(Matrix)。这个100×120矩阵的各个元素主要是1、2、3等简单数据，另外还有某些个体/探针的带型模糊不清或由于其它原因使数据缺失(Missing)，缺失数据用0代替。以上的数据矩阵就是下述连锁分析的基本数据。

对于在两个亲本间只存在一条谱带差异的电泳结果，如同工酶和RAPD等的资料，可将有谱带者赋值为1，谱带缺失者赋值为3。

为了分析主基因控制的农艺性状与遗传标记之间的连锁关系，还必须将有关的表型性状数据化。例如可将群体中的个体按植株的高度分为高秆和矮秆两大类，然后根据亲本的表现分别给高秆和矮秆个体赋值，如1和3。对于群体中不同个体表现为简单二项分布的性状，如植株对病害的反应可分为抗病和感病，对土壤缺磷的反应可分为迟钝（耐性）和敏感等，可将不同类型的个体分别赋以1和3。对中间型的个体则赋值为2。这样经过数据化处理，即可得到各个体与分子标记数据相类似的农艺性状代码值，从而可将有关的性状进行类似于标记的连锁分析。

数据的收集和处理应注意以下问题：(1) 避免利用没有把握的数据：由于分子多态

性分析涉及许多实验步骤，很难避免出现错误，经常会遇到所得试验结果(如X－光片)不清楚等问题。如果硬性地利用这样的数据，不仅会影响该标记自身的定位，而且还会影响到其它标记的定位。因此，应删除没有把握的数据，或重做试验。(2) 注意亲本基因型：如果已知某两个基因座位是连锁的，而所得结果表明二者是独立分配的，这可能是由于亲本类型决定的错误。可试将基因型改变，重新计算。(3) 当两本出现多条带的差异时，应通过共分离分析区别是属于同一基因座位还是分别属于多个基因座位。如属于多个基因座位，应逐带收集分离数据。这是比较不同实验结果的重复性的基础。

4.3.2 连锁的两点测验

当两对基因位于同一染色体上且相距较近时，在分离后代中它们必然表现为连锁遗传。连锁的基本检测是对其分离的成对基因座位进行统计学评价。在进行连锁测验之前，必须了解各检测基因座位的等位基因分离是否符合孟德尔分离比例，这是连锁检验的前提。根据等位基因间显隐性的不同，几种常见群体的孟德尔分离比例如表4.6。

表4.6　不同显性程度下几类群体的孟德尔分离比例

群　体	F_2	BC(三交)	DH (SSD)
共显性	1:2:1	1:1	1:1
显　性	3:1	1:1	1:1

检验RFLP的分离是否偏离孟德尔比例，一般采用 x^2检验(x^2-test)。我们以F_2群体中共显性标记的分离为例加以说明(表4.7)。

表 4.7　共显性下F_2群体孟德尔分离比例的检验

F_2个体带型	P_1 1	Het 2	P_2 3	总计
观测次数(O)	n_1	n_2	n_3	$n=n_1+n_2+n_3$
	33	80	31	144
理论次数(e)	$n/4$	$n/2$	$n/4$	n
	36	72	36	144
$O-e$	-3	8	-5	
$(O-e)^2$	9	64	25	
$(O-e)^2/e$	9/36	64/72	25/36	1.83

在理论上，χ^2 的计算分式为

$$\chi^2=\Sigma(O-e)^2/e$$

上例中

$$\chi^2=(33-36)^2/36+(80-72)^2/72+(31-36)^2/36=1.83$$

所计算出的 χ^2 值与自由度为2的显著水平值 $\chi^2_{0.05(2)}=5.99$ 进行比较。

$\chi^2<\chi^2_{0.05(2)}$ 故该分离比例符合1:2:1。

只有当待检验的两个基因座位各自的分离比例正常时，才可进行这两个座位的连锁检验。目前在分子图谱的制作过程中，发现了大量 RFLP标记偏离孟德分离比例的异常分离(Distorted segregation)。这种异常在远缘杂交组合的分离群体或经一粒传产生的SSD群体中尤为明显。现已证实，存在着一系列影响配子生活力和竞争力的遗传因素。它们的存在导致配子选择。在水稻中已发现了十余个与异常分离有关的基因座位。

异常分离会使连锁的检验受到影响，一些本来不存在连锁的标记由于各自的异常分离，可能导致得出连锁的结论，而另一些本来连锁着的标记也有可能由于异常分离而无法检测到连锁。发生严重异常分离的标记一般不应用于连锁作图。Heun和Gregorius (1987)讨论了异常分离的群体中连锁估计的问题。 Wang等在利用SSD群体构建连锁群时采用了Manly (1994)提出的修正方法。

如果我们同时考虑两个要检验的基因座位 $A-a$ 和 $B-b$，就可以将分离比的检验与连锁检验合并起来进行。这是实际分析过程中常用的方法。下面以相引组合 $P_1(AABB)\times P_2(aabb)$ 的 F_2 为例加以说明。

假定 $A-a$ 与 $B-b$ 的重组率为 r，则 F_2 为分离可预测如表4.8。

根据 $A-a$ 和 $B-b$ 两个座位等位基因的显隐性关系，它们之间的连锁只有三种可能

表 4.8 由 $AABB\times aabb$ 所产生的 F_2 群体中各类基因型的理论比例(重组率为 r)

F_1 配子 频率	AB $(1-r)/2$	Ab $r/2$	aB $r/2$	ab $(1-r)/2$
AB $(1-r)/2$	$AABB$ $(1-r)^2/4$	$AABb$ $r(1-r)/4$	$AaBB$ $r(1-r)/4$	$AaBb$ $(1-r)^2/4$
Ab $r/2$	$AABb$ $r(1-r)/4$	$AAbb$ $r^2/4$	$AaBb$ $r^2/4$	$Aabb$ $r(1-r)/4$
aB $r/2$	$AaBB$ $r(1-r)/4$	$AaBb$ $r^2/4$	$aaBB$ $r^2/4$	$aaBb$ $r(1-r)/4$
ab $(1-r)/2$	$AaBb$ $(1-r)^2/4$	$Aabb$ $r(1-r)/4$	$aaBb$ $r(1-r)/4$	$aabb$ $(1-r)^2/4$

的形式即(1:2:1)-(1:2:1)，(1:2:1)-(3:1)和(3:1)-(3:1)。对于每一种连锁形式，可以进行如下4种 χ^2 测验，即

χ^2_T ：总的卡方值，它可以分解成为三个部分

χ^2_A ：测验 $A-a$ 基因座位 A 和 a 的分离比例是否异常

χ^2_B ：测验 $B-b$ 基因座位 B 和 b 的分离比例是否异常

χ^2_L ：测验 $A-a$ 和 $B-b$ 之间是否存在连锁。

即 $\chi^2_T=\chi^2_A+\chi^2_B+\chi^2_L$

对于不同的连锁形式，χ^2_T 的各个组成部分是不同的。各种情形如下(其中个基因型的字母组合形式用来表示群体中该基因型的个体数)：

(1:2:1)-(1:2:1)

$$\chi^2_T=(4/n)\{AaBb^2+2(AABb^2+AaBB^2+Aabb^2+aaBb^2)+4(AABB^2+AAbb^2+aaBB^2+aabb^2)\}-n \quad df=8$$

$$\chi^2_A=(2/n)\{2(AABB+AABb+AAbb)^2+(AaBB+AaBb+Aabb)^2+2(aaBB+aaBb+aabb)^2\}-n \quad df=2$$

$$\chi^2_B=(2/n)\{2(AABB+AaBB+aaBB)_2+(AABb+AaBb+aaBb)_2+2(AAbb+Aabb+aabb)^2\}-n \quad df=2$$

$$\chi^2_L=\chi^2_T-\chi^2_A-\chi^2_B \quad df=4$$

(3:1)-(1:2:1)

$$\chi^2_T=(8/3n)\{2A-BB^2+A-Bb^2+2A-bb^2+6aaBB^2+3aaBb\ 2+6aabb\ 2\}-n \quad df=5$$

$$\chi^2_A=(4/3n)\{(A-BB+A-Bb+A-bb)^2+3(aaBB+aaBb+aabb)^2\}-n \quad df=1$$

$$\chi^2_B=(2/n)\{2(A-BB+aaBB)^2+(A-Bb+A-bb)^2+2(aaBb+aabb)^2\}-n \quad df=2$$

$$\chi^2_L=\chi^2_T-\chi^2_A-\chi^2_B \quad df=2$$

(3:1)-(3:1)

$$\chi^2_T=\chi^2_A+\chi^2_B+\chi^2_L \quad df=3$$

$$\chi^2_A=(A-B-+A-bb-3aaB--3aabb)^2/3n \quad df=1$$

$$\chi^2_B=(A-B--3A-bb+aaB--3aabb)^2/3n \quad df=1$$

$$\chi^2_L=(A-B--3A-bb-3aaB-+9aabb)^2/9n \quad df=1$$

回交类型 (1:1)-(1:1)

$$\chi^2_T=(4/n)(AaBb^2+Aabb^2+aaBb^2+aabb^2)-n \quad df=3$$

$$\chi^2_A=(AaBb+Aabb-aaBb-aabb)^2/n \quad df=1$$

$$\chi^2_B=(AaBb-Aabb+aaBb-aabb)^2/n \quad df=1$$

$$\chi^2_L=(AaBb-Aabb-aaBb+aabb)^2/n \quad df=1$$

下面举例说明 F_2 群体中(1:2:1)-(1:2:1)连锁及其检验。

$A-a$ (1:2:1) × $B-b$ (1:2:1)

↓

F_2

F_2 自交所产生的各类基因型个体所占比例如下：

	AA	Aa	aa	小 计
BB	27	5	0	32
Bb	6	56	3	65
bb	1	4	30	35
小计	34	65	33	132

$$\chi^2_{\text{总}} = (4/132)[56^2+2(6^2+5^2+4^2+3^2)+4(27^2+1^2+0^2+30^2) -132=165.818$$

$$\chi^2_{A} = (2/132)[2(27+6+1)^2+(5+56+4)^2+2(0+3+30)^2 -132 = 0.045$$

$$\chi^2_{B} = (2/132)[2(27+5+0)^2+(6+56+3)^2+2(1+4+30)^2 -132 = 0.167$$

$$\chi^2_{L} = 165.818 - 0.045 - 0.167 = 165.606 > \chi^2_{0.05(4)}=9.49$$

测验结果表明

$$\chi^2_{\text{总}} > \chi^2_{0.05(8)}=15.5$$

$$\chi^2_{A} < \chi^2_{0.05(2)}=5.99$$

$$\chi^2_{B} < \chi^2_{0.05(2)}=5.99$$

即 $A-a$ 和 $B-b$ 的分离符合孟德尔1:2:1比例。

$$\chi^2_{L} > \chi^2_{0.01(4)}=13.3$$

即 $A-a$ 和 $B-b$ 之间存在连锁。

4.3.3 利用MLE估计重组值

两个连锁座位不同基因型出现的频率是估算重组值的基础。在F_2群体中，$A-a$ 和 $B-b$ 连锁座位的表型频率如表4.9。当 $n=16$，$r=0.5$（$A-a$ 和 $B-b$ 独立）时，很容易计算出上述3类连锁的比例，这时的比例就是 $A-a$ 和 $B-b$ 独立分离时所形成的分离比例。

在一般的遗传学教材中，重组值的估计是根据分离群体中重组型个体占总个体的比例

表 4.9 不同连锁形式下 $A-a$ 和 $B-b$ 两个基因座位的表型频率（各项乘以 $n/4$）

(1:2:1)-(1:2:1)		(3:1)-(1:2:1)		(3:1)-(3:1)	
AABB	$(1-r)^2$	AAB-	$1-r^2$	A-B-	$3-2r+r^2$
AABb	$2r(1-r)$				
AAbb	r^2	AAbb	r^2	A-bb	$2r-r^2$
AaBB	$2r(1-r)$	AaB-	$2(1-r+r^2)$		
AaBb	$2(1-2r+2r^2)$				
Aabb	$2r(1-r)$	Aabb	$2(r-r^2)$		
aaBB	r^2	aaB-	$2r-r^2$	aaB-	$2r-r^2$
aaBb	$2r(1-r)$				
aabb	$(1-r)^2$	aabb	$1-2r+r^2$	aabb	$1-2r+r^2$

来估计的。这种估计方法无法得到估计值的标准误。因而无法对估值进行显著性检验和区间估计。而采用最大似然法估计(Maximum Likelihood Estimation, MLE)方法进行重组率的估计可解决这一问题。下面说明这一方法的基本原理。

在介绍最大似然法之前，我们先介绍有关似然(Likelihood)的概念。

似然定义为特定观测表型出现的可能性，这种可能性与特定的遗传模型相对应。我们假定有一孟德尔基因座位具有 A 和 a 两个等位基因，其中 A 对 a 为显性。则群体中将有三种基因型 AA、Aa 和 aa，两种表现型 $A-$和 aa。当某对夫妻具有基因型 Aa 时，则其子代具有 $A-$表现型的概率为3/4，具有 aa 表现型的概率为1/4。当第一个孩子出生之后，同样的概率对第二个孩子仍然是成立的，因为根据孟德尔定律，一个孩子的表型并不影响另一个孩子的表型表现。换言之，在这个婚配里，子代的表型是相互独立的。因此，如果这对父母有三个孩子，头两个孩子具有 $A-$表型，第三个为 aa 表型，则出现这些表型的总概率——似然率——为(3/4)(3/4)(1/4)=9/64。通俗地说，在64个这样的家庭中，出现上述表现型的家庭有9个。

如果我们假定 $A-$表型出现的概率不是3/4，而是一个不确定的概率 q，则三个孩子具有表型 $A-$，$A-$和 aa 的概率为 $q \cdot q \cdot (1-q)=q^2(1-q)$。为了正规地表示这个问题，我们用 x_i 代表第 i 观测个体的表型。则对于总共作了n次观测，似然值可以写为：

$$L(q)=P(x_1, x_2, \cdots, x_n; q)$$

其中 q 为未知值（称为参数）。如果 n 次观测是相互独立的，$P(x_i; q)$ 代表第 i 次观测的可能性或似然值，则

$$L(q)=P(x_1; q)\,P(x_2; q)\cdots P(x_n; q)=\prod P(x_i; q)$$

在上述三个孩子的实例中，表型 $x_1=x_2=A-$，$x_3=aa$，似然值 $L(q)$ 为

$$L=P(x_1, x_2, x_3)=P(x_1)P(x_2)P(x_3)=q^2(1-q)$$

如果 $q=3/4$，则 $L=9/64$。

现在让我们看看与似然值有关的最大似然法及其在重组率估计中的应用。

最大似然法以满足其估计值在观察结果中出现的概率最大为条件，故又称为最大似然估计。最大化是通过求导并令其为零进行的。

下面以相引组合 $AABB \times aabb$，$A-a$ 和 $B-b$ 均存在显性时，$A-a$ 和 $B-b$ 的连锁为例加以说明。

根据表4.9，F_2分离群体中各个体可按表型分为4组，即 $A-B-$，$A-bb$，$aaB-$和 $aabb$，其理论频率为 $p_i(i=1, 2, 3, 4)$。而 p_i 又是要估计的参数 r 的函数，即 $p_i=f(r)$。其中 $p_1(A-B-)=(3-2r+r^2)/4$，$p_2(A-bb)=p_3(aaB-)=(2r-r^2)/4$；$p_4(aabb)=(1-r)^2/4$。

上述频率 p_i 满足 $\Sigma p_i=1$。

如果令上述4种表型的实际观测次数依次为 n_1，n_2，n_3，n_4，则有 $\Sigma n_i=n$。当以样本容量 n 抽样时，各组观测次数 n_i 的概率分布为多项式 $(p_1+p_2+p_3+p_4)^n$ 的展开。而特定于某一观察的 (n_1, n_2, n_3, n_4) 组合的概率L(r) 为

$$L(r) = \frac{n!}{n_1!n_2!n_3!n_4!}(p_1)^{n1}(p_2)^{n2}(p_3)^{n3}(p_4)^{n4}$$

$$= \frac{n!}{n_1!n_2!n_3!n_4!}(1/4)^n(3-2r+r^2)^{n1}(2r-r^2)^{n2+n3}(1-2r+r^2)^{n4}$$

由上式找出参数 r 的估计值 $\hat{r}$，使它满足 $L(r)$ 为最大，就是对 r 的最大似然估计。这个问题显然只是对方程 $dL(r)/dr=0$ 求根。为便于求微分，可先对上式作对数变换 [$L(r)$ 的自然对数称为Support或log-likelihood]，即

$$\ln L(r) = C + n_1\ln(3-2r+r^2)+(n_2+n_3)\ln(2r-r^2)+n_4\ln(1-2r+r^2)$$

其中 C 为常数，在此

$$C = \ln\frac{n!}{n_1!n_2!n_3!n_4!} - n\ln(1/4)$$

在微分时成为0，可省略，这里求 $\ln L(r)$ 最大值与求 $L(r)$ 最大值显然等价。为求使 $\ln L(r)$ 最大的 r 值，令 $d\ln L(r)/dr=0$。$\ln L(r)$ 的导数称为Score，记为 S，即：

$$S = \frac{-n_1\cdot 2(1-r)}{3-2r+r^2} + (n_2+n_3)\frac{2(1-r)}{(2r-r^2)} - n_4\frac{2(1-r)}{(1-2r+r^2)} = 0$$

即

$$\frac{n_1}{3-2r+r^2} - \frac{n_2+n_3}{2r-r^2} + \frac{n_4}{1-2r+r^2} = 0$$

$$\frac{n_1}{2+(1-r)^2} - \frac{n_2+n_3}{1-(1-r)^2} + \frac{n_4}{(1-r)^2} = 0$$

令 $(1-r)^2 = k$，得

$$\frac{n_1}{2+k} - \frac{n_2+n_3}{1-k} + \frac{n_4}{k} = 0$$

即

$$nk^2+(2n-3n_1-n_4)k-2n_4 = 0 \qquad (n=n_1+n_2+n_3+n_4)$$

$$k = \frac{-(2n-3n_1-n_4)+\sqrt{(2n-3n_1-n_4)^2+8n\,n_4}}{2n}$$

而 $\hat{r} = 1-\sqrt{k}$

这个 $\hat{r}$ 就是重组率r的最大似然估计。根据Rao-Cramer不等式，不难证明，$\hat{r}$ 的抽样方差 $V\hat{r}$ 似近于

$$\frac{1}{V\hat{r}} = -E\left[\frac{d^2[\ln L(r)]}{dr^2}\right] = I$$

上式中的 $\frac{d^2[\ln L(r)]}{dr^2}$ 为 $\ln L(r)$ 对于 r 的二阶导数；E为取期望。其中

$$\frac{d^2[\ln L(r)]}{dr^2} = -\sum_i^k \frac{n_i}{p_i^2}\left(\frac{dp_i}{dr}\right)^2 + \sum_i^k \frac{n_i}{p_i}\left(\frac{d^2p_i}{dr^2}\right)$$

$$E\left(\frac{d^2[\ln L(r)]}{dr^2}\right) = -n\sum_i^k \frac{1}{p_i}\left(\frac{dp_i}{dr}\right)^2 + n\sum_i^k \frac{d^2p_i}{dr^2}$$

$$= n\sum_i^k \frac{1}{p_i}\left(\frac{dp_i}{dr}\right)^2$$

因为 $\sum_i^k \frac{d^2p_i}{dr^2} = \frac{d}{dr}\sum_i^k p_i = 0$

$$\frac{1}{\hat{V_r}} = n\sum_i^k\left[\frac{1}{p_i}\left(\frac{dp_i}{dr}\right)^2\right] = n\sum_i^k i_i = I$$

上式中 I 称为总信息量，$\sum i_i = I/n$ 称为单一观察信息量。

根据上述公式，可按表4.10 求 r 的方差。

$$I = n\left[\frac{(1-r)^2}{3-2r+r^2} + \frac{2(1-r)^2}{2r-r^2} + 1\right] = \frac{2n(3-3r+2r^2)}{(3-2r+r^2)(2r-r^2)}$$

$$\hat{V_r} = \frac{(3-2r+r^2)(2r-r^2)}{2n(3-3r+2r^2)}$$

由此可见，n 越大，$\hat{V_r}$越小，估计越精确。

将表4.10 的 n_i 值代入求 k 公式得

$$k = \frac{1927 \pm \sqrt{1927^2+8\times6952\times1338}}{2\times6952} = 0.7743$$

$$r = 1 - \sqrt{k} = 1 - \sqrt{0.7743} = 0.1201$$

$$\hat{V_r} = 1.76702\times10^{-5}$$

表 4.10　两对连锁基因完全显性时F_2群体重组率估计值的方差计算

组别	n_i	期望频率 p_i	dp_i/dr	$i_i = \frac{1}{p_i}\left(\frac{dp_i}{dr}\right)^2$
A−B−	4831	$(3-2r+r^2)/4$	$-2(1-r)/4$	$i_1 = \frac{(1-r)^2}{3-2r+r^2}$
A−bb	390	$(2r-r^2)/4$	$2(1-r)/4$	$i_2 = \frac{(1-r)^2}{2r-r^2}$
aaB−	393	$(2r-r^2)/4$	$2(1-r)/4$	$i_3 = \frac{(1-r)^2}{2r-r^2}$
aabb	1338	$(1-r^2)/4$	$-2(1-r)/4$	$i_4 = \frac{4(1-r)^2}{4(1-r)^2} = 1$
总和	$6952=n$	1	0	$\sum i_i = \frac{(1-r)^2}{3-2r+r^2} + \frac{2(1-r)^2}{2r-r^2} + 1$

重组率 $\hat{r}$ 为 $0.1201 \pm \sqrt{1.76702 \times 10^{-5}} = 12.01\% \pm 0.42\%$

以上是 $AABB \times aabb$ 产生的 F_2 群体中，表现为显性的一对基因间重组率及其标准误的计算。对于其它群体且等位基因间表现不同显性的情形， Allard (1956)给出了完整的求 $\hat{r}$ 和 $V_{\hat{r}}$ 的计算公式。只要将相应的公式代入即可进行类似的计算。

4.3.4 似然比与连锁检验

在人类遗传学研究中，由于通常不知道父母的基因型或父母中标记基因的连锁相是相斥还是相引，因而无法通过计算重组体出现的频率来进行连锁分析。所以人类的连锁分析通常采用似然比 (Likelihood ratio or odds ratio)的方法(Fisher 1935, Haldane和Smith 1947, Morton 1955)。这就是比较观测资料来自某一假设(如两个标记间以 r 的重组率相连锁)的概率与来自另一备择假设(通常为非连锁 $r = 1/2$)的概率。这两种概率之比为一种假设相对于另一假设的似然比，即 $L(r) / L(1/2)$。在上述 $A-a$ 和 $B-b$ 两对基因均存在等位基因显性作用的例子中，

$$L(r) = \frac{n!}{n_1! n_2! n_3! n_4!} (1/4)^n (3-2r+r^2)^{n_1} (2r-r^2)^{n_2+n_3} (1-2r-r^2)^{n_4}$$

$$L(1/2) = \frac{n!}{n_1! n_2! n_3! n_4!} (1/4)^n (2.25)^{n_1} (0.75)^{n_2+n_3} (0.25)^{n_4}$$

为了计算上的方便，常将 $L(r) / L(1/2)$ 取以10为底的对数，称为 LOD 值，

$$LOD = \log_{10} \frac{L(r)}{L(1/2)}$$

当 $n=6952$， $n_1=4831$， $n_2=390$， $n_3=393$， $n_4=1338$ 时，有

r	0.05	0.10	0.15	0.20	0.25	0.30	0.12
LOD	586.42	682.51	678.52	632.01	560.79	472.54	688.04

对于不同的 r 值，LOD 值是不同的，且当 $r=0.12$ 时，LOD 值最大。一般而言，当 $A-a$ 和 $B-b$ 连锁时，$L(r)/L(0.5)>1$，则相应的 LOD 为正。而当 $L(r)/L(1/2)$ 小于1时，相应的 LOD 值为负。注意 $L(r)$ 可能为0，在特定的参数值下，本来这种情形是不可能发生的但碰巧发生了。例如在连锁分析中，从重组本该发生的数据中可能得出似然值为0的估值。此时 LOD 将为负无穷大。这种 LOD 值的解释就是观测资料中应该存在一次重组事件。

为了证实两对基因之间存在连锁，在人类遗传学中一般要求似然比 $L(r) / L(1/2)$ 大于1000:1，即 $LOD>3$，才能证实这两对基因间存在连锁，而要否定连锁的存在，只要求似然比大于100:1，即 $LOD<-2$。

在其它生物遗传图谱的构建中，似然比的概念也用来反映重组率估值的可靠性程度或作为连锁是否真实存在的一种判断尺度。

4.3.5 多点分析与基因直线排序

上述介绍的连锁测验及重组率估算方法都是建立在两点分析的基础上的，即每次只考虑两个标记基因座位之间的连锁关系。在遗传图谱的构建中，两点分析只是一个起点。由于两点分析只利用了有限数目的标记共有信息，所估计的重组率只是一个大概值。要从这样的重组率去推测基因的排列顺序可能导致错误的结论。同时考虑多个标记基因座位的共分离，将可以获得有关这些座位排列顺序的可靠信息。

上面介绍的最大似然法及与它有关的似然比法，可以用于各种标记座位排列顺序的比较，以便从多种可能图谱中挑选出具有最大可能性的图谱，并计算出这些基因座位之间的重组率。下面我们以三点分析为例介绍和引入有关的方法。

如果考虑 A、B、C 三个基因座位，每两个座位之间的重组率可假定为 r_{AB}，r_{AC} 和 r_{BC}。从组合 $AABBCC \times aabbcc$ 中产生的 F_2 群体，可以依照组合 $AABB \times aabb$ 的方式，推算出 F_2 群体基因型的种类及其在特定 r_{AB}、r_{AC} 和 r_{BC} 下各种基因型出现的频率。我们可以采用前面介绍的同样方式对涉及 r_{AB}、r_{AC}、r_{BC} 的似然函数 $L(r_{AB}, r_{AC}, r_{BC})$ 求导，并令其为零。但我们无法采用同样的方式得到各重组率的最大似然估计值，因为此时要估计的参数有 r_{AB}、r_{AC} 和 r_{BC} 三个，而可用的估计式只有一个。为了得到最大似然估计值 r_{AB}、r_{AC} 和 r_{BC}，必须采用一种常用的数值分析法——迭代法 (Iteration method)。

对于多点分析 (Mutiple analysis)，假定 $M_1, \cdots, M_m$ 为 m 个遗传标记座位，并按正确或假定的染色体顺序排列。对于 m 个基因座位，一共有 $m!/2$ 种可能的基因排列顺序。采用计算机可以同时分析多种基因排列顺序，从而构建一个最佳的连锁图谱。假定两个毗邻基因座位 M_i 和 M_{i+1} 之间的重组率为 r_i，即

$$M_1 \quad M_2 \quad M_3 \quad M_4 \quad M_5 \quad \cdots \quad M_{m-1} \quad M_m$$
$$r_1 \quad r_2 \quad r_3 \quad r_4 \quad \cdots \quad r_{m-1}$$

我们的目的是要找到 $r=(r_1, r_2, \cdots, r_{m-1})$ 使观测资料产生的机会最大。为了简单起见，先不考虑交叉干扰 (Crossover interference)。

设两个毗邻基因座位 M_i 和 M_{i+1} 之间的重组率为 r_i，即要找到 $r=(r_1, r_2, \cdots, r_{m-1})$ 使观测资料产生的可能性最大。其似然函数为

$$L \propto p_1(r_1, r_2, \cdots, r_{m-1})^{n1} \cdot p_2(r_1, r_2, \cdots, r_{m-1})^{n2} \cdot \cdots \cdot p_n(r_1, r_2, \cdots, r_{m-1})^{nn}$$

取自然对数之后，分别对 r_1，r_2，……，r_{m-1} 求偏导，并令其为0，得到一组方程

$$\ln L / r_1 = \cdots \cdots = 0$$
$$\ln L / r_2 = \cdots \cdots = 0$$
$$\cdots$$
$$\ln L / r_{m-1} = \cdots \cdots = 0$$

求 r 需要在一个多维空间进行寻找。必须定义一种迭代方法产生修正的试估值 r^{new} 来代替旧的试估值 r^{old}，使似然值增加。人类遗传学中常用的迭代方法是EM算法 (Dempster等1977)。这种迭代法可用于处理属于不完整的观测资料，换言之，存在不唯一确定重组与非重组体数目的资料。EM算法每一迭代步骤包括一次求期望 (Expectation，E)

和随后的一个最大化(Maximization, M)。而在实验生物学中，遗传学家可以得到完整的数据——即发生在每一区间(M_i, M_{i+1})的重组的和非重组体的数据。在完整数据下，最大似然值r_i可以通过计算重组体数目来确定:r_i就是重组体数占总个体数的比例。EM算法为从不完整资料获得最大似然估计值提供了十分有效的方法。这种方法在连锁分析中的步骤如下:

(1) 给定一个试估值$r^{old}=(r_1, r_2, \cdots, r_{m-1})$

(2) 就象该试估算为真实重组值一样，用于计算完整资料的期望值，即在每一区间重组体和非重组体的期望数。

(3) 把这种为完整资料所求的期望值当作真值，计算重组值的最大似然估计值r^{new}。

(4) 迭代(2)、(3)两个步骤，直到似然值收敛于最大值为止。

EM算法仅为进行E和M步骤提出了特定的程序，必须根据其具体应用使算法具体化。对于连锁作图，M步骤很简单:$(r^{new})_i$就是第i区间减数分裂重组体的期望数与总的减数分裂数目的比值。困难在于E步骤。这步我们称为“遗传重建问题”:即给定重组率r，计算每一次减数分裂中发生重组体减数分裂的期望数。Lander和Green(1987)讨论了遗传重建的几种方法。传统的方法每次迭代需要m次似然值计算，而他们所讨论的方法中，只需2次传统的似然值计算。其中以Hidden Markov链重建法为最佳，其计算时间不象其它方法随基因座位数目的增加呈指数增加，而是呈线性增加，从而大大缩短了计算时间，使采用大量标记进行同时多点分析成为可能。

Lander等(1987)编制了电子计算程序Mapmaker，采用Marhov重建法进行遗传重建，以获得最大似然的遗传连锁图。当对数似然值(Log-likelihood)上下迭代间的增量低于某一给定的Tolerance值T时，就认为该步骤收敛了。Lander和Green(1987)证明似然值不会随迭代而减少，因而只要选用一个正的Tolerance值就可以了。在实际运算中，常选的T值为0.01或0.001。

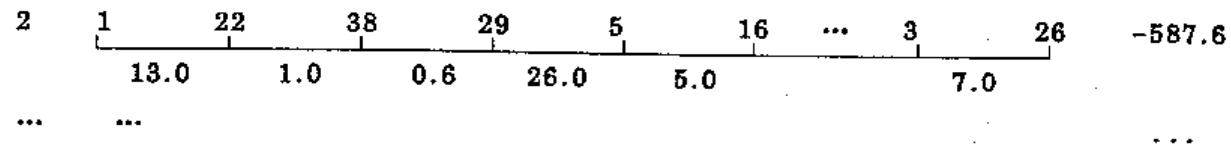
Lander和Green(1987)给出了一个采用EM算法进行多点连锁分析的实例。在所研究的人类CEPH(The Centre d' Etude de Polymorhisme Humaine)家系中，第7染色体上有16个标记，共构成15个区间。开始时，给定这些区间重组值的初估值r_i均为0.05(即紧密连锁)。自此计算的相应对数似然值为-391.45。当T=0.01时，经过12次迭代后，各区间重组值收敛于某一特定值(T=0.01)。此时相应的对数似然值为-303.28。所以收敛时所得图谱产生观测资料的可能性比开始迭代时假定r_i=0.05时的可能性高出$10^{-303.28-(-391.45)}=10^{48}$倍(表4.11)。

对于$m!/2$种可能的基因排列顺序，都可以得到一个各自最大的log-likelihood。比较这些可能的排序，log-likelihood最大者即为最佳的顺序。例如:

排序		log-likelihood
1	1 22 38 29 5 16 … 8 26	-590.0
	13.0 1.0 0.6 26.0 5.0 8.0	

表 4.11　利用EM法进行多点连锁分析的实例（引自Lander 和Green 1987）

迭代	重组率															log(*L*)
0	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	-351.45
2	0.01	0.05	0.18	0.09	0.13	0.10	0.08	0.09	0.08	0.08	0.14	0.09	0.09	0.08	0.21	-306.68
4	0.01	0.04	0.22	0.08	0.16	0.11	0.08	0.10	0.08	0.08	0.18	0.08	0.08	0.06	0.24	-304.25
6	0.01	0.03	0.24	0.07	0.17	0.11	0.08	0.11	0.08	0.08	0.20	0.08	0.08	0.06	0.25	-303.66
8	0.01	0.03	0.25	0.06	0.18	0.11	0.07	0.11	0.08	0.08	0.22	0.08	0.08	0.05	0.25	-303.43
10	0.01	0.03	0.25	0.06	0.18	0.11	0.07	0.11	0.08	0.07	0.22	0.07	0.08	0.05	0.25	-303.34
12	0.01	0.03	0.25	0.05	0.19	0.11	0.07	0.11	0.08	0.07	0.23	0.07	0.08	0.05	0.25	-303.28



生物的染色体数是特定的，位于同一染色体上的分子标记将连锁在一起形成一个连锁群。按照上述多点作图的原理，同一染色体上的标记联系在一起的可能性将大于其它染色体上的标记。先对所有标记进行初步分群，然后在每一群内寻找可能性最大的排序，最终可获得所有连锁群的标记排序，即各染色体的分子遗传图谱。我们在水稻分子遗传图谱的构建中，以籼稻窄叶青8号和粳稻京系17为亲本，杂交产生F_1，并自交产生F_2作图群体，采用Mapmaker作图软件进行了分子遗传图谱的构建(徐云碧等1994b)。美国 Tanksley 实验室现已报道了较为饱和的分子标记连锁图谱(Causse等 1994)，图4.3 给出了其中的第6染色体。

4.3.6　交换干扰与作图函数 (Mapping function)

随着两基因间距离的增加，染色体上两个基因座位之间便可能在两处同时发生遗传物质的交换即双交换。由于双交换的结果，两个基因间的两线双交换就不能被鉴别出来，从而使根据重组值估计的两个基因座位间的图距估计值偏低。

对于 $A-B-C$ 三个连锁基因，令 r_1 和 r_2 分别为 $A-B$ 和 $B-C$ 间发生单交换的概率，那么，若两个单交换的发生彼此独立，则在 $A-B$ 和 $B-C$ 之间同时发生交换即形成一次双交换的概率为 $r_1 \times r_2$ 。事实上，在染色体有关区段上发生双交换的实际次数往往少于由单交换概率相乘所得的理论值。这种倾向表明，染色体上一个区段内所发生的交换可以减少相邻区段内交换的发生。这种现象称为染色体干扰(即交叉干扰)简称干扰。

干扰可以分为两种类型，即染色单体干扰和交叉干扰。干扰的程度可用符合系数(Coefficient of Coincidence) C 表示，符合系数 C 为实际双交换值对理论双交换值的比值。理论双交换值是指所研究的两个区段彼此独立时可能产生的双交换值。

$$C = \frac{\text{实际双交换值}}{\text{理论双交换值}} = \frac{\text{实际双交换值}}{r_1 \ r_2 \cdot n}$$

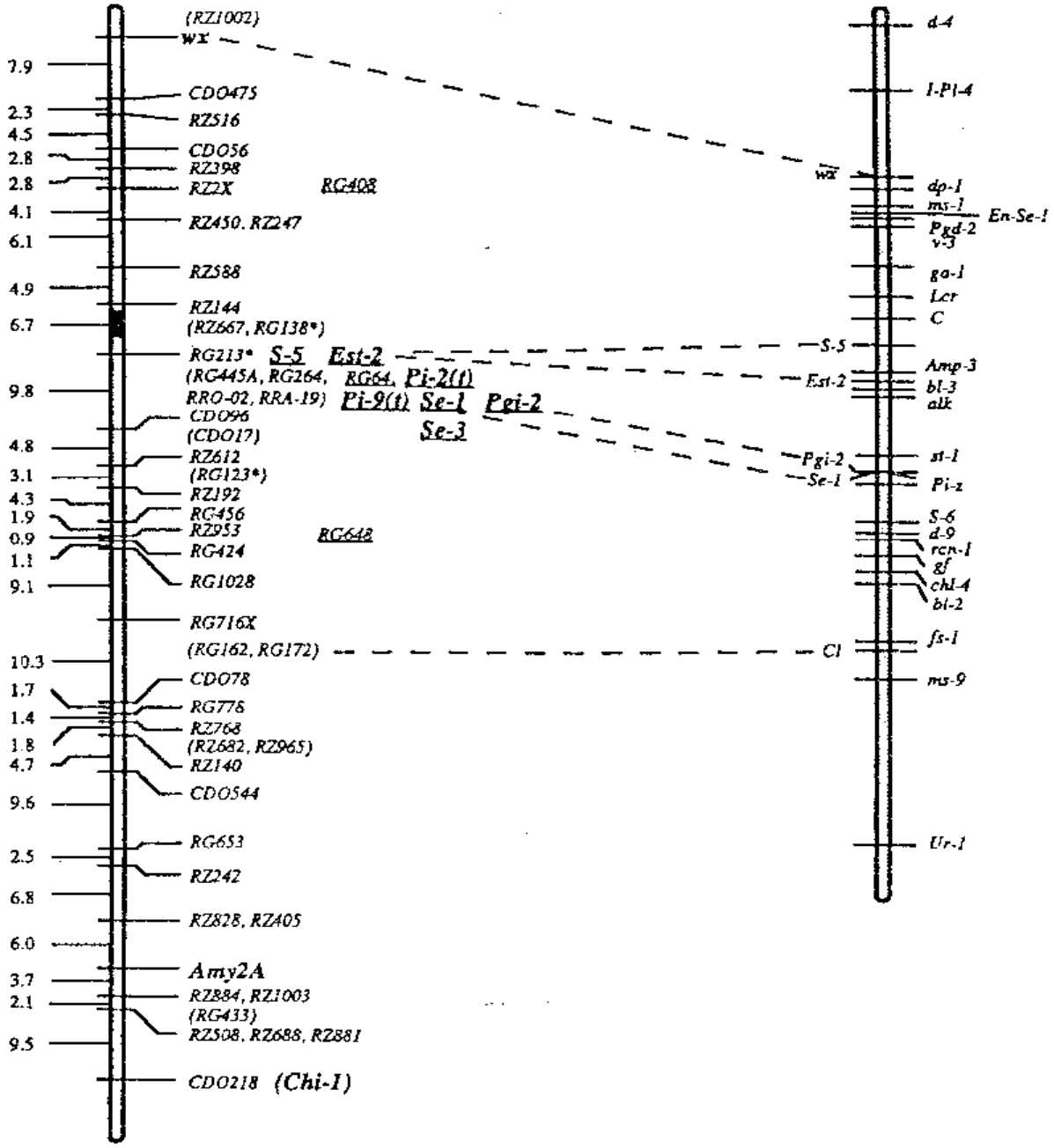


图 4.3 水稻第6染色体的分子标记连锁图谱(左)和经典遗传图谱(右)

r_1 和 r_2 分别为两个相邻染色体区段发生单交换的概率。n 为所统计的后代总数(包括重组体和非重组体)。C 的标准误可由下式计算

$$S_C = \pm\sqrt{(C/n)\cdot[(1-C\,r_1-C\,r_2-C\,r_1r_2+2\,C^2r_1r_2)/(r_1\,r_2)]}$$

符合系数 C 值变动于0～1之间。当 C=0时，表示发生完全干扰，没有双交换发生，这意味着所研究的相邻染色体区段一定很短。当 C=1时，表示两个区段没有干扰，单交换可单独发生。意味着两个染色体区段相对较长。

当没有中间基因可供利用时，要测量两个相距较远的基因间的图距时，有必要采用一些数学方法进行矫正以降低多重交换对交换值(或重组值)估计的不利影响，这样就无需从表现型(或基因型)上鉴别由多重交换产生的个体的实际数量。这种矫正可为真实图距(cM)与重组率(%)提供精确的函数关系。故将这种矫正函数称之为作图或定位函数(Mapping function)。

如果不对某染色体区域内所发生的双交换作矫正，则以重组率估计的图距就会比真实图距低2 $C\,r_1\cdot r_2$，若以 r_s 表示两区段之间总的重组率，则它们的关系可用下式表示：

$$r_s=r_1+r_2-2\,C(r_1\cdot r_2)$$

当完全干扰时，C=0，于是 $r_s=r_1+r_2$，当完全不发生干扰时 C=1，有

$$r_s=r_1+r_2-(r_1\cdot r_2)。$$

在 C=1假定下，Haldane(1919)给出了两个基因座位之间的真实图距 x 与重组值 r 之间的关系：

$$x=-(1/2)\ \ln(1-2r)$$

其中 x 为作图距离，以交换单位、图距单位或cM来评价。利用Haldane的公式，22%的重组率相当于图距为 $-(1/2)\ \ln(1-2\times0.22)=0.29$，即29cM。

Haldane作图函数的不合理之处在于假定了完全不发生交换干扰。Kosambi(1944)在不作此假定的情况下推导出如下的作图函数：

$$x=(1/4)\ln\frac{1+2r}{1-2r}$$

当 r=0.22时，x=23.6cM。这一作图函数在遗传学研究中得到更为广泛的应用。

4.3.7 染色体着丝粒和端粒的定位

一个完整的染色体具有几个主要部分：着丝粒(Centromere)，缢痕(Constriction)和随体(Satellite)及端粒(Telomere)。这些基本结构在生物染色体的运动与复制等方面起着重要的作用。此外，有些组成结构还是遗传图谱制作中不可忽视的部分。要得到一个完整的遗传图谱，必须知道染色体上的标记与着丝粒之间的距离。由于着丝粒并不是一个基因，不能从表型测知，故采用常规的两点、三点乃至多点分析方法是无法确定标记与着丝粒之间的关系的。

在经典的遗传图谱构建中，一般采用近端着丝粒染色体来对基因与着丝粒之间的距离进行定位。近端着丝粒染色体是正常染色体在着丝粒附近断裂形成的异常染色体。在小麦中，目前已获得全部42条染色体的近端着丝粒染色体。以下介绍怎样用近端着丝粒染色体来确定基因与着丝粒之间的距离。

将带有隐性基因aa的近端着丝点染色体个体作母体，而正常染色体上带有显性基因AA的个体作父本，二者杂交后，得到一条带A基因的正常染色体和一条带a基因的近端着丝粒染色体的杂合体，将此杂合体作父本再与携带aa双隐性基因的母本回交。如果未发生交换，则产生图4.4中①和②二种基因型的后代，如果发生交换，则新产生的后代基因型为图中的③和④。由于带有近端着丝粒的花粉无法与正常花粉竞争，故在上述实验中，只形成没有发生交换作用的后代Aa(图中的①)和A与着丝粒发生过交换的后代aa(图中的④)。将发生交换的后代数目除以总后代数，就得到基因A-a与着丝粒之间的相对距离。

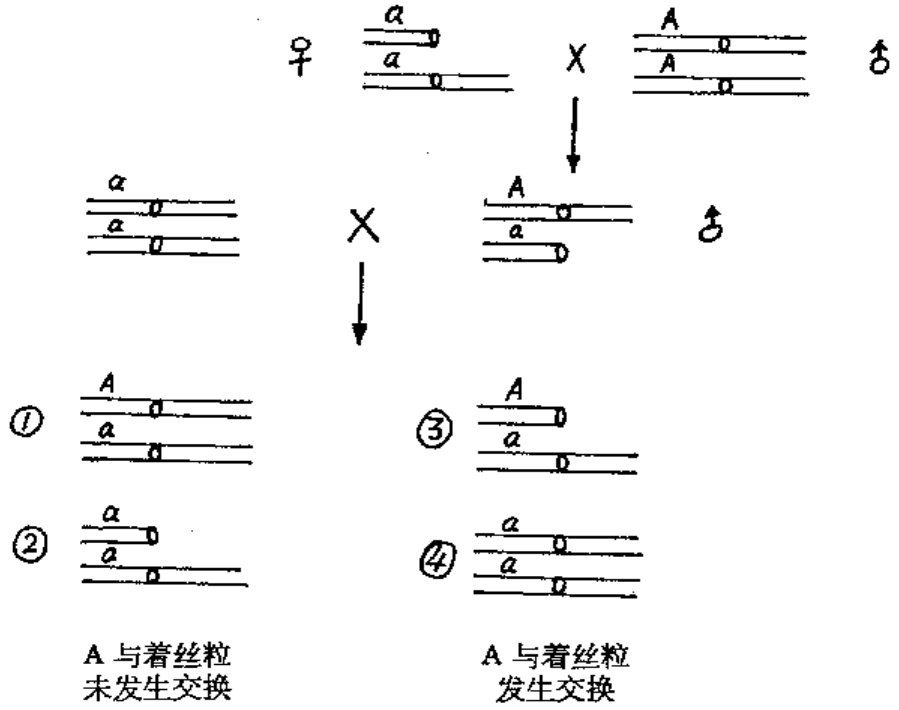


图 4.4　利用近着丝粒染色体进行着丝粒定位

也可以利用染色体易位材料来判断着丝粒在染色体上的位置。一般易位点和着丝粒所在部分的交换被抑制，因而推算位于着丝粒两旁的易位点与标记基因间的重组率时一般都偏小。利用这个现象可以推算连锁图上着丝粒的位置。在细胞学上，利用已知易位点的易位系统进行基因分析也可知道着丝粒的位置。早在1945年，在玉米中就利用易位分析的结果推测了全部染色体的着丝粒位置。

染色体上的端粒(Telomere)是指染色体的自然末端。端粒具有独特的行为特征。正常情况下，它们不会在染色体结构变化中被插入到染色体的中间位置。端粒缺失时，染色体的正常行为通常会受到破坏，因为染色体的断裂末端将会与其它类似的末端相连接。

在遗传图谱的构建中，端粒位置的确定就意味着为染色体的全长设定界标。对低等真核生物端粒所作的研究表明，端粒是高度保守的，并表现以下特征:(1)它们含有串连重复的，短的，C丰富的简单序列块；(2)由于串连重复块在数量上存在很大变异因而导致端粒大小的不同；(3) 许多端粒 DNA 双链中具有单链不连续性；(4)它们不被包装在核小体

(Necleosome) 中；(5) 它们可能具有末端回折结构 (Foldback structure)。由于传统凝胶电泳方法分辨能力的限制，在大多数情况下无法将具有多态性的端粒片段区分开来，因此RFLP分析之类的程序不能简单地用于端粒的作图。在分子杂交时一般要借助于脉冲凝胶电泳 (Pulsed-Field Gel Electrophoresis, PFGE)将有差异的片段分离开来。

已从几种高等真核生物中克隆到了位于染色体端部的与端粒有关的序列。其中，四膜虫(*Tetrahymena*)的线性染色体外rDNA 的端粒是第一个被测序确定的端粒。随后尖毛虫 (*Qxytricha*)、锥虫 (*Trypanosoma*)、酵母、拟南芥、番茄以及人的染色体端粒区DNA 先后被克隆、测序。端粒区在不同生物中大小不同，从20bp到几个kb不等。即使在同种生物甚至同一生物个体细胞的不同染色体上，端粒区的DNA长度都是各不相同的，这主要是由于端粒重复序列重复的次数不同造成的。由于端粒DNA序列的重复性和保守性，采用一般的分子作图方法难以确定端粒与染色体其它部位基因的距离。

Richards和 Ausubel (1988) 发展了一种富集端粒DNA序列的基因组文库构建法，使得可以从具有较大染色体的高等真核生物分离出端粒，并将这一方法应用于克隆拟南芥的端粒DNA序列。最近Burr等(1992)在玉米染色体上找到了三个特殊的限制性片段与端粒的存在正相关。在番茄中已定位了几个串联重复的端粒DNA序列(Ganal等1992)。利用PFGE与切割位点稀少的限制性酶相结合，Wu和Tanksley (1993) 研究了水稻端粒结构的特征，采用来自拟南芥的端粒探针将三个水稻的端粒带分别定位在第8、9、11染色体上，并证实了多态性的端粒片段不仅在遗传上而且也在物理上与遗传图谱上最远端的RFLP标记相连锁。寻找连锁群末端的方法还将涉及到把遗传作图与染色体的原位杂交相结合，以确定标记在遗传和细胞学图谱上的实际物理位置。

4.3.8 分子标记连锁群的染色体定位

构建了一个比较饱和的分子图谱之后，下一步的工作是把分子标记所建立的连锁群与经典遗传图谱联系起来，从而将其与特定的染色体挂钩。完成这一任务，可以根据分子标记与已知染色体位置的形态标记的连锁来进行。此外，在水稻中还可以利用非整倍体，在玉米中利用A/B易位系、在小麦中利用零体/四体染色体代换系等，将分子连锁群与特定的染色体联系起来。

以水稻为例，n=12。目前已获得到各个染色体的全套初级三体。初级三体的染色体数为2n+1，在水稻某一三体的总DNA中，由于该染色体有一式3份，这条染色体DNA的含量为其它11条染色体DNA含量的1.5倍。在DNA定量相当准确的条件下，用已知能检测某一连锁群的探针分别与12种三体的DNA杂交，由于杂交强弱与同源顺序的含量成正比，即存在剂量效应，因此根据杂交后的每种DNA滤膜放射自显影显带的强弱，就可以看到一种三体的强度明显高于其它11种，说明这个标记所对应的顺序就在份数为三的这条染色体上。

此外，当原位分子杂交的灵敏度达到可以揭示单拷贝序列的杂交位点时，就可采用原位分子杂交将连锁群的分子标记定位到染色体上。

4.3.9 经典遗传图谱与分子遗传图谱的综合

从1987年报道玉米和番茄的RFLP遗传图谱以来，具有重要经济价值的栽培植物几乎全

部都已构建了以RFLP等分子标记为主的遗传图谱，其中番茄和水稻分子图谱上所定位的标记数已达到了1000个以上。为了充分利用目前可供利用的分子和遗传的信息，必须将分子遗传图谱与经典的遗传图谱结合起来，使其成为一个可以相互联系两者的综合遗传图谱。

但是，由于两类图谱的构建是相互独立的，它们之间很少共有的遗传标记，两类图谱在构建时采用了不同的作图群体和不同的统计分析方法。经典遗传图谱的构建是不同的实验室、利用不同的作图群体、在不同的实验条件下完成的，通常一次只研究少数遗传标记，在统计分析上一般不考虑重组率估计的标准误，也不对双交换进行校正，因而标记所在的位置和不同标记间的相对位置可能是值得怀疑的。将两类遗传标记综合到一个遗传图谱中去，不仅是重要经济性状准确定位的需要，也是以遗传/物理图谱为基础，选择适当的探针分离目的基因的需要。

在栽培植物中，番茄和水稻的经典连锁图谱和分子连锁图谱的综合进展较快。Weide等(1993)构建了番茄第6染色体的综合图谱，确定了分子标记与10个经典遗传标记的连锁关系。水稻综合图谱的构建受益于在经典连锁图谱上的长期累积性工作，目前已将47个表型标记，11个同工酶标记定位到了分子连锁图谱上(Causse等1994)。

4.4 用于构建遗传图谱的计算机程序包

遗传图谱的构造需要对大量标记之间的连锁关系进行统计分析。随着标记数目的增加，计算工作量常常成指数形式增加。过去经典图谱的建立是在一段很长的时期内逐渐补充和完善的，不同标记之间连锁分析的资料来源于不同的时间、不同的群体及不同的研究者，因而无法利用所有与标记有关的资料所能提供的最大信息。而分子图谱构建所用资料通常是在短时间内由某一研究小组从单一分离群体中获得的。为了充分利用所有资料的信息，需要对全部标记同时进行分析。这种分析涉及十分庞杂和大量的计算，是手工无法完成的。借助于现代计算机，可以对大量数据进行随心所欲地分析和处理。许多学者为构建遗传图谱设计了一些专用程序包。概括起来，有以下几种：

(1) LINKAGE (Lathrop和Labouel 1984)：用于人类连锁分析(多点分析)

(2) Mapmaker (Lander和Botstein 1989)：用于人类、植物近交物种间F_2，BC群体

(3) Mapmaker/EXP (Lincoln 等1993)：可用于RIL群体

(4) MAPL (Y. Ukai 私人通信)：小型计算机可以使用(Pascal设计)，适合F_2等群体

(5) Map Manager (Manly 1993)：可用于RIL群体。

上述专用程序以Mapmaker在植物分子连锁图谱构建中应用最为广泛。尽管不同版本的同一软件在操作上可能有很大的差别，但一些常用的命令则改变不大。下面介绍Mapmaker在人类CEPH型资料和F_2型资料连锁作图中的应用。我们将以程序的操作过程为顺序，介绍其主要的操作命令及其功能。所有命令都跟在 ">" 符号之后。

4.4.1 人类CEPH型数据的连锁作图

1. 启动程序(Starting Mapmaker)

>load data (调用数据文件)

>photo(产生一个可供打印时调用的文件)

>sequence *all (选用所有标记来排序)

>two point (两点分析)

计算所有成对座位的重组率和最大LOD值，推测可能的连锁群，一般要求LOD≥4.0，r=0.30 。

2. 建立小连锁群 (Mapping a small linkage group)

>sequence { } 利用多点测验将某染色体上的标记进行各种可能的排序。

>compare 计算每一种最大的似然图谱，并按likelihood的大小排列出来，打出order，及log-likelihood值

3. 图谱显示(Displaying a map)

>sequence 将上述分析结果给定一个新的 "sequence" 进行分析。

>map 显示cM图距和重组率及该图的likelihood 。

>let 给这个 "sequence" 一个名称，如 chrom1=1234 。

4. 建立一个更大的连锁群(Mapping a larger linkage group)

当标记较多时无法采用上述方法进行排序。解决的办法是在连锁群的一个亚群里进行所有可能的多点分析。在此之前，先利用前面所得 "two-point" 数据。

>big lods 2.0 将LOD值≥2.0的所有成对座位显示出来。

根据以上信息，选择一组标记进行多点分析，相间为10-20cM的标记较容易排序。

5. 将未定位标记加入已有连锁群

4.4.2 F_2 数据的连锁作图

1. 启动 Mapmaker

>load data

>photo

2. 推测可能连锁群

>sequence *all 或指定某一顺序

Mapmaker将只利用 "sequence" 最新指定的顺序(order)进行分析。

>two point 对所有成对的座位进行两点测验，计算其重组值和最大LOD值并将信息贮存备用。

>group 0.4 3.0 根据两点分析结果，推测可能的连锁群。只有当某一标记基因座位与该 "group" 中的另一标记座位的最大LOD值≥3.0，重组率≤0.4时，该标记座位才并入到该 "group" 中。计算机将显示可能的连锁群(suspected linkage groups)。

3. 利用三点测验构建连锁群

先利用三点测验确定各 "group" 的可能顺序。然后利用n点分析确定这些"order" 的最大似然图距。

>three point 0.4 3.0 计算所有标记座位的各个三连体的三点图谱及相应的 "likelihood"。然后利用三点测验资料确定每一 "group" 的可能 "order"。最后利用n点测验计算这些 "order" 的最大似然图距和相应的likelihood。推断连锁群要求r≤0.4，

LOD≥3.0。

Mapmaker将按图谱 "likelihood" 的大小排列所有可能的三连体，并显示该序顺相对于最可能顺序之间的相对ln(L)。例如对于三连体 {1 29 5}，有

likelihood 排序 {1 29 5} > {29 1 5} > {1 5 29}

相对log-likelihoods -0.00 -3.47 -11.81

最佳图谱为

$$1 \underset{20.5cM}{\text{———}} 29 \underset{35.1cM}{\text{———}} 5$$

完成所有三点图谱之后，就可以确定各连锁群中那一些 "order" 不与三点资料发生严重冲突。当某三个标记座位的三点顺序的可能性比另一顺序的可能性高1000倍时，即将此顺序列入Mapmaker。

在计算了每一连锁的所有可能 "order" 之后，Mapmaker将利用多点分析(full n-point analysis)计算这些 "order" 的最大似然图距及 "likelihood"。当某连锁群有一个以上的 "order" 时，可以采用多点似然差异来排除各种可能的顺序(order)。

继续上述分析，完成其余连锁群的排序。

>let 采用此命名给连锁群的某 "order" 以染色体编号。

4. 构建连锁群的其它途径

仅利用 "full multipoint analysis"，也可以构建连锁群。

>sequence 定义待分析的 "sequence"，采用{ }。

>compare计算连锁群中的每种可能顺序的最大似然图谱及相应的似然值"likelihood"。然后比较似然值，并选择最可能的 "order"。Mapmaker将根据 "likelihood" 的大小，最多挑选20个种较佳的 "order"。

5. 显示连锁图

>sequence *chrom2等。

>map

"map"引导Mapmaker计算当前 "sequence" 定义的每一 "order" 的最大似然图谱。如果当前的 "sequence" 定义了一个以上的 "order"，如 {6, 10, 23, 28, 29, 40} 共有360种可能的 "order"，则显示所有特定的图谱。

>print names on

>map

Mapmaker含有每一标记座位的名称，采用 "print names on" 之后再 "map"，所显示的图谱中将以标记座位名称代替编号。

6. 将新的标记座位增加到已定位的顺序中去

>sequence

>near

为了将某一标记座位增加到 "sequence" 定义的 "order" 中去，"near" 将计算新的标记座位与目前 "order" 中每一个标记之间的两点距离及最大LOD 值。输入 "near"命令后，Mapmaker将要求输入判断连锁所需的最大重组值和最小LOD 值。并显示新的标记座位与 "order"各座位间的重组值及LOD 值。由此可以推断新的标记座位所靠近(near) 的座位。

>try

"try" 将新的标记座位依次放在每个可能的两个标记座位之间，计算出所产生"order"的最大似然图谱。通过比较所有可能的图谱，便可确定新座位最可能的存在位置(对于每一种似然图谱，"try" 都将重新计算所有可能的重组率)。

对于每个标记座位区间，Mapmaker将显示支持新的标记座位插入该区间的"相对log-likelihood"。该值是将新的标记座位插入该区间所得似然值与新的标记座位插入其最可能的区间所得似然值之间的差值。因此该值为零即表明了最佳的插入位置。

全部结果贮存在photo所产生的文件中，可供打印时调用。

7. 结束

采用quit命令。

4.5 与遗传图谱构建有关的几个问题

4.5.1 构建饱和图谱所需基因座位数

遗传图谱所需的饱和度是指单位长度染色体上已定位的标记数或标记在染色体上的间隔距离。一个基本的染色体连锁框架图谱大概要求其上的标记平均间隔不大于20cM。为了利用连锁图谱进行主基因的定位，其平均间隔还要求更小一些如10～20cM。对于数量性状的基因定位，标记的平均间隔要求在10cM以下，才能最大可能地找到的所有有关的QTL。为了满足反向遗传学进行基因克隆、染色体步行等的需要，要求目标区域标记图谱的平均间隔在1cM以下。

不同生物基因组的大小有极大差异，因此满足上述要求所需的定位标记数是不同的。以人类和水稻为例，它们的基因组全长分别为3.3×10^9bp和5.8×10^8bp，如要构建一个平均图距为0.5cM的分子图谱，则所要定位的标记数就要分别达到6600和3200个。几种生物不同图谱饱和度下所需定位的标记数如表4.12所示。

表 4.12 遗传图谱达到特定饱和度(标记间平均距离)所应包含的标记数目

生物		人类	水稻	玉米	拟南芥	番茄
基因组大小	(kb)	3.3×10^6	5.8×10^5	7.2×10^6	7.0×10^4	7.1×10^5
	(cM)	3300	1600	1550	500	1500
	bp/cM	1000	362.5	4645	140	473
图谱饱和度	20cM	165	80	76	25	75
	10cM	330	160	155	50	150
	5cM	660	320	310	100	300
	1cM	3300	1600	1550	500	1500
	0.5cM	6600	3200	3100	1000	3000

Tanksley等(1988)给出了所需标记数与图谱饱和度之间的关系。以12条染色体，每条长100cM，全长1200cM的生物为例。有两个因素影响着所需标记数的大小。(1)标记间的平均距离，即总的图谱单位长度除以定位的标记数。对于染色体全长1200cM的生物，如果定位了120个标记，则标记间的平均距离为10cM。(2)标记间的最大距离，这也许比平均距离更为重要。因为它决定基因组所有部分被覆盖的程度。图4.5对这一变量的估计是根据泊松定理(99%概率水平)进行的，并假定重组值是可加的。从图4.5中可以看出，当平均间隔为1cM时，而图上两个标记间的最大可能间隔为10cM。所以尽管大部分区域标记的间隔为1cM或更小，但仍有可能找到间隔10cM的空隙(gap)。大于10cM的空隙的概率小于0.01。为了将最大可能间隔从10cM减少到5cM。将需要另外增加1000个标记。在实际上，通过随

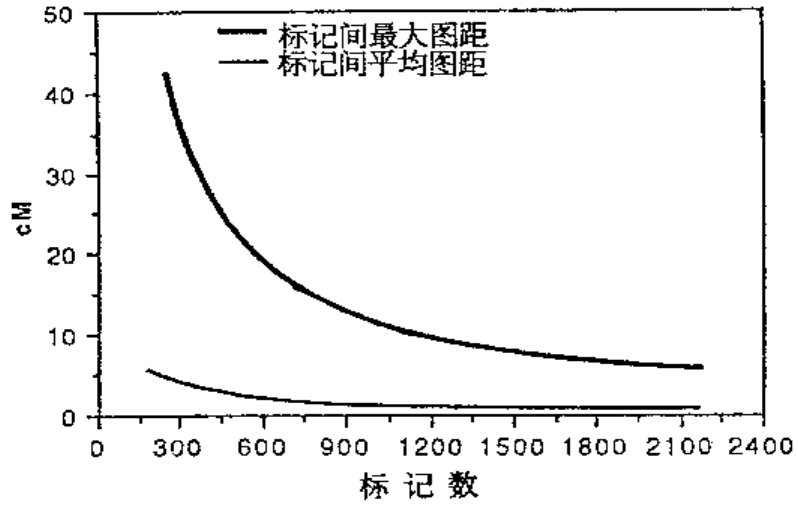


图4.5 遗传图谱所含标记数对平均图距和最大图距的影响
(假定基因组全长1200cM，最大图距的置信限为(99%)

机作图方法产生间隔均一且为1cM或更小的图谱是不实际的。为了填补空缺，有必要选择可以填补这些空缺的克隆作标记探针或寻找该区域上有差异的亲本构建作图群体。

单拷贝序列在染色体分布上的不一致性，以及染色体上重组事件发生的异质性，也可能导致遗传图谱中出现较大的空缺。大量的证据表明，单拷贝和重复DNA序列在染色体上的分布不一定是均一相隔分布的。端粒和着丝粒通常具有更大比例的重复DNA，因而单拷贝克隆可能不足以代表这些染色体的区域。从这一点考虑，利用不同分子标记特性的互补将有助于减小遗传图谱中的空隙。

4.5.2 群体或组合数目

遗传连锁作图的工作量很大，同时利用几个组合是不现实的。现已发表的RFLP连锁图谱大多是基于一个组合，仅有个别情况涉及到两个不同组合。然而，纵观已有的RFLP连锁图谱，或多或少地都有基因座位分布不均匀的问题。在基因组的某些区域基因座位相对集

中，而在另一些区域基因座位又相对稀少。造成这种结果有两种可能的原因：一是染色体不同部位发生交换的机率不等。一般认为着丝粒附近的交换频率较低，而染色体两端的交换率较高。因此，如果所用探针在物理图谱 (Physical map) 上是均匀分布的话，在遗传图谱上则会表现为着丝粒附近标记密集，染色体两端标记稀少；二是双亲在这些染色体区段不存在多态性。后一种情形的解决办法就是要选用其它组合。另外，连锁图谱的构建是手段而不是目的。要用其解决遗传育种中的不同问题，就需要有相应的组合。再则，多配制几个组合的工作量有限，这不仅可以为以后的图谱构建提供选择，而且还可以避免试验过程中等待试验材料的问题。

在植物基因组研究中，有时需要多个亲缘关系差异不同的作图群体或组合。对于饱和作图，利用两个差异很大但相互亲和的亚种(或生态型)甚至栽培种×野生种产生F_2、BC或RIL群体是十分有益的。在这类F_2中，可以预期大多数DNA探针可以比在栽培品种间检测到更高水平的多态性。这样可以定位大量的探针，从而增加遗传图谱的饱和度。我们建议把这作为栽培植物完整作图的第一步（表4.13)。在远缘杂交(如栽培品种×野生种)的组合中更容易鉴别和定位分子标记 。因此采用两个差异极大的材料可以获得特定物种的饱和分子图谱。尽管与其它类型的标记相比，一个物种内栽培品种间存在较大的分子多态性，但要发现分子水平的差异仍然需要检测大量的探针/酶组合。也就是说，并不是所有在物种水平上可以揭示多态性的探针都可在品种间比较中揭示多态性。因此，要利用根据育种目标所选择的两个特定栽培品种构建饱和的分子图谱是比较困难的。但是高度饱和的物种图谱是构建品种间图谱必需的第一步。为此首先要根据探针在物种图谱上的位置，从中选择代表目标基因组区域的探针。然后用这些探针来检测品种间的多态性(表4.13,第二步)。品种比较中未能检测到多态性的探针可用同一区域的其它探针来代替，直到在目标区域找到多态性探针为止。这样就从总的物种探针中找出了可以适当区分这两个栽培品种的一组探针。这种为特定的品种－品种比较揭示分子水平多态性的两步程序，其优点在于：不必为每对栽培品种检测和定位大量的随机DNA探针。

4.5.3 群体大小

估算两个基因座位间的遗传距离以多大的群体为宜？这取决于如下几个因素：1) 群体类型（如F_2，回交等)；2) 基因座位间的连锁状态（相引或相斥)；3) 性状的分离情况（如9∶3∶3∶1或15∶1等)；4) 基因座位间的距离。分子标记一般是共显性遗传，是最简单的一种分离情况。因此相对于形态性状的遗传分析，构建分子连锁图谱所需群体较小。现已发表的连锁图谱所用分离群体一般不超过100个单株或家系。但如利用这样大小的群体去定位那些控制农艺性状的基因，尤其是数量性状基因，所得试验误差就会很大。如果建图的目的是用于基因组的序列分析或基因分离等工作，更需用较大的群体，以保证所建连锁图谱的精确性。因此我们认为每个群体都应保证有足够的个体（如500个以上)。为减小工作量，骨架连锁图的构建可基于大群体中的一个随机小群体(如150个单株或家系)。当需要精细地研究某个连锁区域时，再有针对性地在骨架连锁图的基础上扩大群体。这样大小群体相结合，既可达到研究目的又可减轻工作量。

如果从作图效率考虑，作图群体所需样本容量的大小决定于以下两个方面：一是从随

表4.13　利用不同作图群体进行两步作图的战略

第一步：物种图谱	
	—— 利用差异极大但相互亲和的品系或种间组合进行分子作图
	—— 构建高度饱和的物种图谱
优点：	每一探针均可揭示差异
	只需要使用有限的限制性内切酶
结果：	获得高分辨率图谱
	对该物种具有普遍的利用价值
问题：	并非总是可以直接用于每一特定的栽培品种间组合
	详细遗传图距可能因组合而变化
第二步：品种－品种图谱	
	—— 利用物种图谱中已定位探针确定多态性的分子标记
	—— 构建特定的分子图谱，定位未覆盖的区域
	—— 从这些区域选择探针，并用于寻找紧密连锁的分子标记
优点：	可以非随机性地分离片段特异的分子标记
	无需进一步作图
	可产生有用的单倍型标记
	可拓展物种图谱

机分离结果可以辨别的最大图距，二是两个标记间可以检测到重组的最小图距。因此，作图群体的大小根据研究者的试验目的确定时，作图群体越大，可以检测到重组的最小图距越小，可以辨别的最大图距越大，即可以将标记定位得更为紧密，同时也可发现更为松弛的连锁。例如在样本容量为100的群体中，当发现一个重组体时，其重组率为1%≈1cM，而在 n =50的群体中，一个重组体所导致的重组值为2%≈2cM，当 n =1000时，一个重组体导致的重组值为0.1%≈0.1cM。

对于可以辨别的最大图距 Max，$Max = r + t_{0.01,n-2}(SE_r) < 0.5$cM。其中 n 为群体大小，t 为 $df = n-2$ 的1%显著水平时的 t 值，r 为重组值的点估计值。

根据以上讨论，不同群体下可以发现的最大图距以及可以细分的最小图距如图4.6所示。

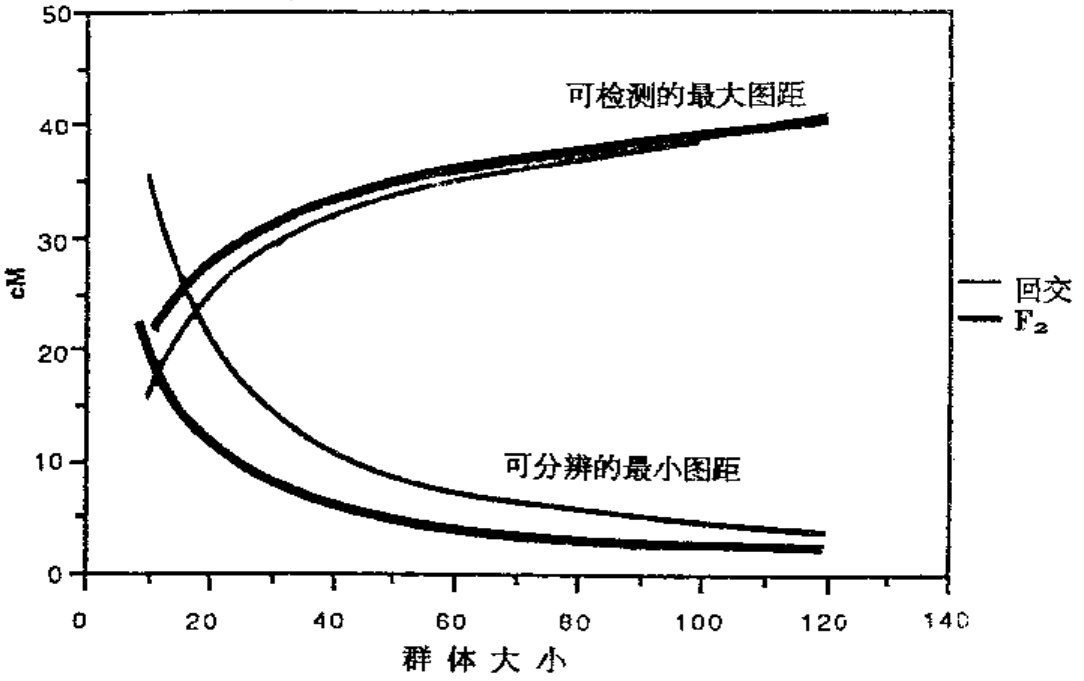


图4.6 F_2和BC群体的大小对可以检测到的最大图距和可以分辨的最小图距的影响
(曲线的置信限为99%)

第 5 章　群体的遗传结构

可用于构建分子连锁图谱的生物群体是多种多样的。它们在图谱构建及以图谱为基础的遗传育种研究中各有其特点。本章将着重介绍几种永久性遗传群体及其在分子作图中的应用。

5.1　群体的性质及其分类

5.1.1　根据群体遗传组成划分

在遗传学上，根据等位基因是否相同，可把基因型分成两类：(1) 纯合的 (Homozygous)，即等位基因相同的，如AA；这种基因型的自交后裔不会分离，除非发生突变。(2) 杂合的 (Heterozgyous)，即等位基因不同的，如Aa；这种基因型的有性后裔必发生分离，从而产生由遗传因素引起的变异。

一个群体，可由单一或若干基因型的个体所组成。如果由单一基因型组成，则不论该基因型纯合或杂合，该群体即为同质的 (Homogeneous)，即不同个体间基因型是一致的；如果由若干不同的基因型组成，则该群体称为异质的 (Heterogeneous)，即不同个体间基因型是异质的。

根据以上定义，常见的植物群体可归类为：

(1) 同质纯合群体，如自交作物的品种，异交作物的自交系，其群体中各个体的基因型均为AA；

(2) 同质杂合群体，如自交作物的F_1，异交作物自交系间F_1，无性繁殖系，其群体中各个体的基因型均为Aa；

(3) 异质纯合群体，如自交作物杂交后的高世代品系所组成的群体和地方品种组成的群体，如 RIL (SSD)、DH等，群体中各品系的基因型是不同的，如AA、aa等，而品系内个体间是相同的，均为AA或aa；

(4) 异质杂合群体，如自交作物的杂种分离世代或异交作物的不同品种，其群体中不同个体或品种的基因型是不同的，既有杂合体，也有纯合体，如AA、Aa、aa。

5.1.2　根据群体的持续性划分

根据群体能否保持其特性连续多代不变，可将其划分为：

(1) 暂时性群体 (Tentative population)：这类群体中分离单位是个体，包括自交作物的分离世代，异交作物的分离世代，如F_2、F_3、F_4、BC、三交群体等。群体一经自交或近交就会使个体的遗传组成发生变化，无法永久使用。

(2) 永久性群体 (Permanent population)：这类群体的分离单位是品系，不同品系之间存在基因型的差异，而品系内个体间的基因型是相同的，自交不分离，如RIL(SSD)、

DH等。这类群体可通过自交或近交繁殖后代，而不会改变群体中各品系的遗传组成，可以永久使用。

5.1.3 根据群体遗传背景划分

(1) 遗传背景同质的近等基因系群体(Near-isogenic line population)：这类群体由供体亲本反复与轮回亲本回交产生的一系列具有大致相同遗传背景的品系组成，品系内不存在个体间基因型的差异，而品系间除目标性状外遗传背景基本一致即其它性状差异很小。

(2) 遗传背景杂合的群体：除近等基因系外的其它遗传群体，如F_2、BC、RIL、DH等，这些群体中个体或品系间所研究的目标性状和其它性状均存在差异，因此目标性状的研究是在异质的遗传背景下进行的。

5.1.4 根据群体来源划分

根据群体中不同个体(品系)的来源，可以将其划分为：

1. 自然品种群体：由一系列品种组成，根据目标性状进行选择并划分为一定的类群。通过研究不同目标性状的品种类群间其它性状的差异，探讨目标性状与其它性状的关系，如通过高秆和矮秆两类作物品种的比较研究探讨株高的遗传效应等。

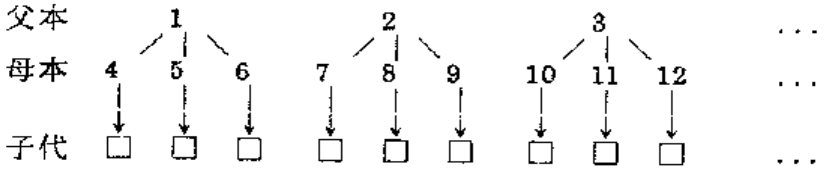
2. 遗传交配群体(Genetic mating populations)：根据遗传研究需要设计的遗传群体，可选择不同的遗传材料按特定的交配方式杂交产生。

(1) 双列杂交 (Diallel cross)：以n个品种(系)分别作父本和母本，进行各种可能的杂交，利用全部F_1或F_2进行遗传分析。交配方式如下(×表示杂交)：

亲本	P_1	P_2	P_3	...	P_n
P_1	×	×	×	...	×
P_2	×	×	×	...	×
...	...			...	
P_n	×	×	×	×	×

(2) NC设计 (North Carolina design)，共有三种：NCⅠ、NCⅡ、NCⅢ

NCⅠ：由两个亲本P_1和P_2杂交并自交产生F_2群体，然后从F_2中选取不同的个体作父、母本相互杂交(每个父本分别与多个不同的母本杂交)产生子代群体：



NCⅡ：将n个亲本分成两组(n_1和n_2)，进行两组品种(系)间的各种可能的杂交：

品种	1	2	3	4	...	n_1
n_1+1	×	×	×	×	...	×
n_1+2	×	×	×	×	...	×
n_1+3	×	×	×	×	...	×
...						
n_1+n_2	×	×	×	×	...	×

NCⅢ：从F_2群体选择n个体分别与父本(P_1)和母本(P_2)杂交(回交)，即

F_2	1	2	3	4	...	n
P_1	×	×	×	×	...	×
P_2	×	×	×	×	...	×

(3) 三重测交(Triple Test Cross, TTC)和简化三重测交(Simplified Triple Test Cross, sTTC)

TTC：由P_1和P_2杂交产生F_1，并由F_1自交产生F_2，然后从F_2群体中选择n个体(n一般大于20)分别与父本(P_1)、母本(P_2)及F_1杂交：

F_2	1	2	3	4	...	n
P_1	×	×	×	×	...	×
P_2	×	×	×	×	...	×
F_1	×	×	×	×	...	×

sTTC：从品种(系)的自然群体中选择n个材料(n一般大于20)作亲本，分别与自然群体中目标性状表现极端的两类品种(系)杂交，其中P_L和P_S分别为自然群体中目标性状表现值最大和最小的品种(系)。

亲本	P_1	P_2	P_3	P_4	...	P_n
P_L	×	×	×	×	...	×
P_S	×	×	×	×	...	×

以上这些遗传交配所产生的后代群体通常用作经典数量遗传学的研究群体。

3. 近交群体

F_1经自交或同胞交配所产生的F_2、F_3等分离群体以及F_1与某一亲本短期回交产生的BC群体或经长期回交产生的NIL群体等。

以下主要介绍永久性群体DH、RIL (SSD)和NIL及其在分子作图中的应用。

5.2 DH 群体

高等植物的单倍体(Haploid)是含有配子染色体数目的孢子体。来自二倍体物种的单倍体叫单元单倍体 (Monoploid)，而来自多倍体物种的单倍体可称为多元单倍体(Polyhaploid)。单倍体经过染色体加倍形成的二倍体称为加倍单倍体或双单倍体(Doubled haploid)。

5.2.1 DH系的产生途径

概括单倍体产生的途径，可分为以下几种：

1. 染色体消失 (Chromosome elimination)

在栽培大麦(*Hordeum vulgare* L.)与球茎大麦(*H. bulbosum*)的杂种F_1中，发现了染色体选择性消失导致形成单倍体的现象。已知影响单倍体形成的因素包括：基因型，生长发育温度(高温消失率高)、亲本基因组比例等。在大麦上，这是产生DH的主要途径。

2. 花粉培养(Anther culture)或雄核发育 (Androgenesis)

取F_1植株的花粉进行离体培养，诱导产生单倍体植株。有很多因素影响花粉植株的诱导，包括：供体植株的生理状况，花药预处理，花粉小孢子发育时期，培养基的组成，培养条件(光照、温度、湿度等)。目前存在的问题主要是培养过程中可能存在基因型和培养基的选择性、高白苗率以及培养期间产生的变异(非亲本间差异所产生的变异)。在小麦和水稻中，花粉培养是产生单倍体的主要途径，单倍体经自然加倍或经秋水仙素处理加倍成为双单倍体。

3. 子房培养 (Ovary culture) 或雄核发育 (Gynogenesis)

4. 单倍体起动基因 (Haploid initiation gene, *hap*)

在大麦中，有报道表明当带有*hap*基因的纯合体作母本与其它品种杂交时，8%的后代是单倍体的，而反交则不产生单倍体。

5.2.2 DH系的评价

1. 配子随机性

产生DH的系统不应对某特定的配子具有选择性。大麦中通过球茎大麦染色体消失途径所产生的DH系一般是随机的，不存在异常分离。Park等(1976)和Choo等(1982)通过比较DH和SSD系统未发现DH系统的配子选择性。但在水稻中，特别是在远缘杂种后代经花培产生的DH群体中，发现了与同工酶，RFLP以及形态性状等有关的异常分离(如单基因座位下两种纯合体的分离偏离期望的1:1比例)。

2. 遗传稳定性

在理论上，DH系具有两个特点，品系内的完全同质性(Homogeneity)和完全纯合性(Homozygosity)。除了培养过程中可能引起的变异之外，DH一旦稳定之后，具有上述两个特征的DH系并不存在“娇娇者易污”或容易检测到突变发生的问题，而是与常规育成的品种一样稳定。目前对培养过程中可能产生的变异还知之甚少。

3. 无性系变异 (Somaclonal variation)

组织培养过程中产生无性系变异的原因可归结于：总的核型变化，隐微染色体重排，体细胞交换和姊妹染色单体交换，转座因子，基因扩增(Amplification)和缩减(Diminution)。这些在组织培养过程中产生的无性系变异也存在于组织培养产生单倍体植株的过程中，并且因此与DH群体中可能出现的染色体数目和结构的差异、等位基因的产生与消失、杂合体的存在、基因的异常分离等有关。

5.2.3. DH群体的数量遗传学

如果DH是来自F_1杂种的配子随机样本，则可用于数量遗传研究。与二倍体遗传模型不同，DH遗传模型排除了显性效应，不需要作不存在上位性之类的不现实的假设，因而可以研究控制数量性状的基因加性上位性及连锁。DH系可以在多个环境或季节进行重复试验，是研究基因型和环境互作的理想材料。DH的加性遗传方差总是大于相应的二倍体系统，因而可以获得较好的选择响应。Choo等(1985)详细讨论了与DH群体有关的数量遗传学，内容包括：基因互作的检测，遗传方差的估计，连锁的检测，基因数目的估计，多基因定位，遗传理论和假设的验证。我们在此不作详细论述。

5.2.4 DH群体在分子作图中的应用

设两个亲本P_1和P_2的基因型分别为 *AABB* 和 *aabb*。其F_1(*AaBb*) 产生4种类型的配子 *AB*、*Ab*、*aB* 和 *ab*，经单性生殖产生4种类型的单倍体植株。加倍后形成 *AABB*、*AAbb*、*aaBB* 和 *aabb* 四种类型的二倍体，当 *A-a* 和 *B-b* 独立时其比例为1：1：1：1 (图5.1)。某中 *AB* 和 *ab* 为亲型配子，*Ab* 和 *aB* 为重组型配子，*AABB* 和 *aabb* 为亲型个体，而 *AAbb* 和 *aaBB* 为重组型个体。

P_1 (*AABB*) × P_2 (*aabb*)

↓

F_1 (*AaBb*)

配　子		*AB*	*Ab*	*aB*	*ab*
单倍体		*AB*	*Ab*	*aB*	*ab*
二倍体		*AABB*	*AAbb*	*aaBB*	*aabb*
比例：	独立	1/4	1/4	1/4	1/4
	连锁	$(1-r)/2$	$r/2$	$r/2$	$(1-r)/2$

图 5.1　DH群体中两个基因座位的分离

对于单个分子标记的分析，DH群体中在理论上只应出现两类亲本的带型。亲本、F_1及DH群体分子标记的带型如下：

P_1	P_2	F_1	DH							
			1	2	3	4	5	6	…	n
—		—	—		—	—	—			
	—	—		—				—		—

下面以水稻和大麦为例说明DH群体在分子标记连锁作图中的应用。

在水稻中，McCouch等（私人通信）以从籼型水稻Apura和粳型旱稻Irat177的F_1经花药培养产生的66个DH系为材料，探讨了利用该群体构建RFLP图谱的可能性。35个RFLP标记的分析结果表明，在DH群体中存在部分杂合性（两个亲本带型的探针如RG227），非亲本等位基因（如探针RG349，图5.2），某些基因座位的异常分离（35个RFLP标记有10个显著偏离1:1）。

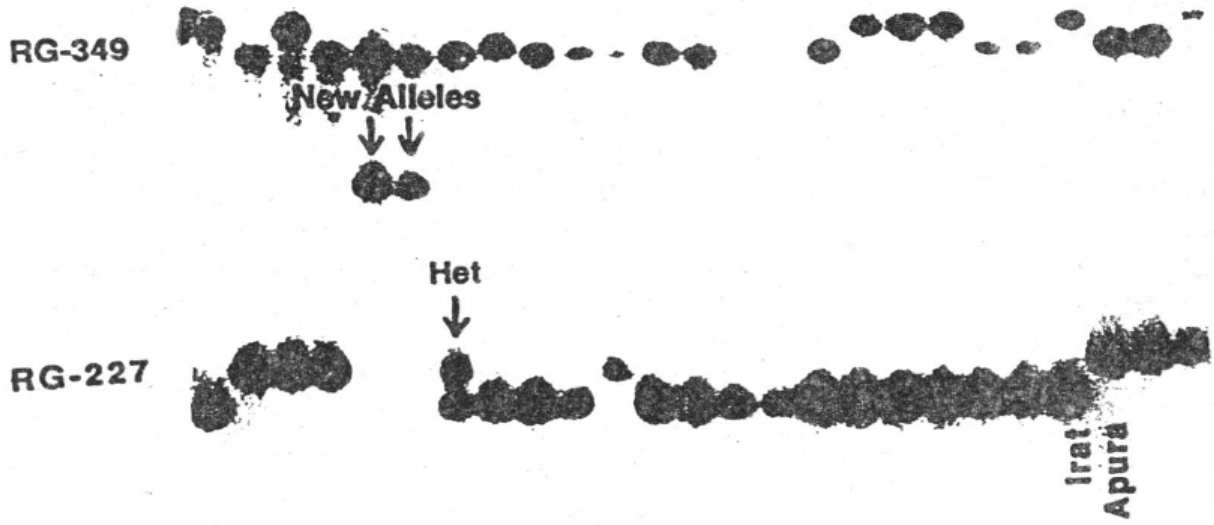


图 5.2 水稻Apura/Irat 177 DH群体中出现的非亲本等位基因（New Alleles）及杂合体（Het）

McCouch等认为该DH群体用于分子作图的真正限制不在于上述因素，而是因为双亲间的多态性较小。在检测的100个探针中，只有约40%发现多态性。定位于IR34583×Bulu Dalam F_2中的克隆只有55%在该DH群体双亲间揭示出多态性。认为双亲可能由于遗传的结果含有一部分共同的基因组，因而不适合于构建饱和的遗传图谱。

Heun等（1991）采用春大麦品种Prottor和Nudinka之间的F_1花药培养产生的113个DH系为作图群体，构建了一张由155个RFLP标记及两个已知基因组成的分子图谱。这是在农作物中利用DH为作图群体完成连锁图谱构建的第一例报道。在157个定位标记中，有15个分离显著偏离1:1，这些显著偏离的标记集中分布在1S、2S、5L和6L等4个染色体臂上。位于异常分离的标记附近的其它标记也表现出某种程度的偏离（表7.1）。

这种异常分离标记在染色体上分布的特异性表明，可能存在导致配子选择性消失的遗

传因素。 在水稻中利用经典的遗传标记所进行的研究证实了导致配子选择性消失或存活力差异的10余个基因座位。 徐云碧和申宗坦 (1992a) 对水稻籼粳杂交组合的研究，发现了水稻胚乳糯质基因的异常分离，与3:1的理论比值相比，糯与非糯的比例可以表现为大于3:1，等于3:1和小于3:1，糯性籽粒所占比例为8.9%～95.6%，可以大大偏离25%的理论范围。 我们在利用籼粳F_2群体所进行的RFLP分析也揭示出异常分离标记在染色体上分布的特异性(徐云碧等(1995d)。由此推断，作图群体中存在的分子标记的异常分离很可能由于遗传原因所造成。

表 7.1 大麦DH群体中异常分离的分子标记在染色体上的分布

染色体 1S		染色体 2S		染色体 5L		染色体 6L	
标 记	%P	标 记	%P	标 记	%P	标 记	%P
Xcnl.CDO545	41	Xcnl.GB140B	61	Xcnl.WG241	24**	Xcnl.BCD221A	58
Xcnl.CDO420A	44	Xksu.DG D18	64**	XksuDG E11	28**	Xcnl.BCD339B	62
Xcnl.BCD129	38	Xcnl.CDO57	64**	Xcnl.CDO393	26**	Xucr.Dhn3,4	62
Xcnl.BCD130	36**	Xcnl.BCD175	64**	Xcnl.CDO394B	45	Xcnl.WG222B	60
Xcnl.WG834	33**	Xcnl.WG516	62	Xksu.DG E2B	36**	Xcnl.CDO419A	66**
Xcnl.WG789	37**			Xcnl.CDO989	40	Xcnl.BCD269	65**
Xcnl.CDO475	39			Xcnl.BCD265	43	Xcnl.WG282	68**
Xcnl.CDO36	44			Xcnl.BCD304	43	Xcnl.BG140A	68**
Xcnl.CDO348	46					Xksu.DG F2	61
Xcnl.CDO771B	34**						

%P —— 亲本 Proctor 等位基因的频率 (%)

5.3 RIL (SSD) 群体

5.3.1 近交及其遗传效应

重组近交系 (Recombinant inbred lines) 或随机近交系 (Random inbredlines) 是将F_2个体连续自交或同胞交配，或群体内随机交配直到完全纯合所产生的一系列品系。

在遗传学上，近交有两个作用，一是使杂种的基因得以分离并进行重组，二是使分离的基因型逐代趋于纯合。例如从一对基因的杂合体 *Aa* 开始，一代接一代地进行自交，就一代一代地分离为 *AA* 、*aa* 纯合体及 *Aa* 杂合体。随着自交世代的进展，群体中纯合体所占比例逐代增大，杂合体逐代减少。其过程如下：

在一对基因的情形下，每自交一代，纯合代数目将增加一半，杂合体数目将减少一半。在自交第 t 代，杂合体所占比例为 $(1/2)^t$， 两种纯合体所占比例为$1-(1/2)^t$，其中AA和aa纯合体各占一半，即 $[1-(1/2)^t]/2=(2^t-1)/2^{t+1}$ (表7.2)。

当两对或两对以上独立遗传基因的杂合体进行连续自交时， 例如 k对基因，在自交第 t 代，杂合体所占的比例为 $(1/2)^{tk}$，纯合体所占比例为 $[1-(1/2)^t]^k=[(2^t-1)/2^t]^k$。

表 7.2　一对基因杂合体自交所产生的基因型及其频率

世代	基因型			杂合体频率	纯合体频率 (%)
	AA	Aa	aa		
0	-	1	-	1	0
1	1/4	2/4	1/4	1/2	50.0
2	3/8	2/8	3/8	1/4	75.0
3	7/16	2/16	7/16	1/8	87.5
4	15/32	2/32	15/32	1/16	93.8
5	31/64	2/64	31/64	1/32	96.9
10	1023/2048	2/2048	1023/2048	1/1024	99.9
...					
t	$(2^t-1)/2^{t+1}$	$2/2^{t+1}=1/2^t$	$(2^t-1)/2^{t+1}$	$1/2^t$	$1-1/2^t$

两个亲本有差异的基因对数越多，则其杂种在同一自交世代群体中纯合体所占的比例就越少，即从杂合体转化为纯合体的速率就越慢(图5.3)。例如同为自交7代，某初始杂合体为

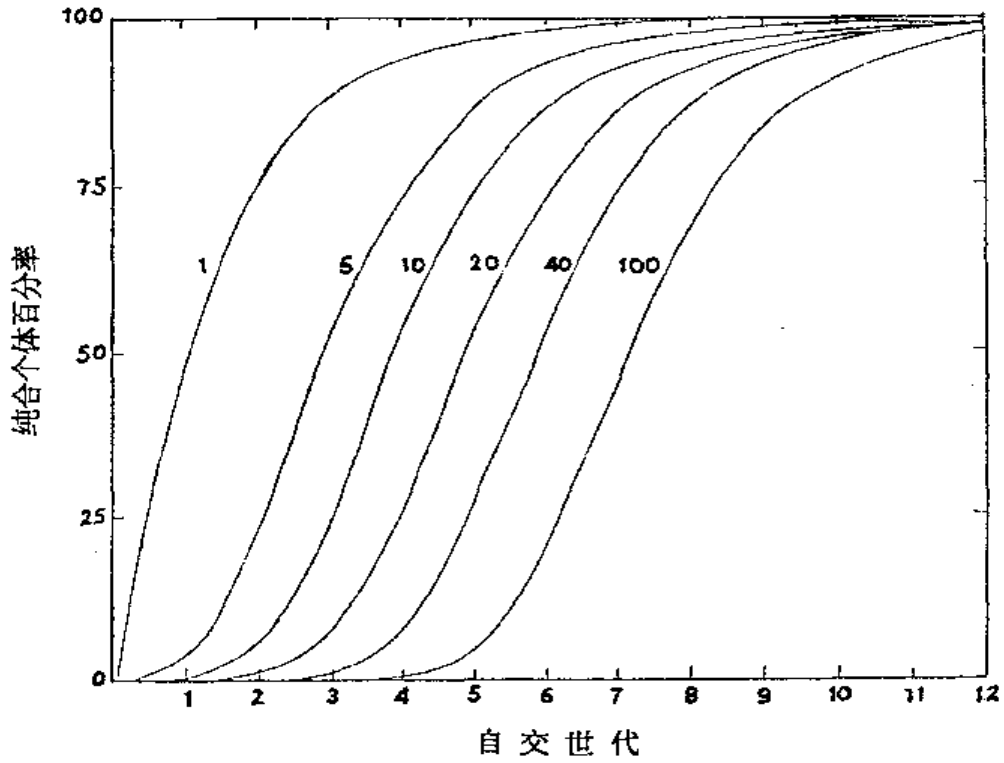


图5.3　自交世代和杂合基因数对自交群体纯合体频率的影响
(自交世代数分别为1、5、10、20、40、100)

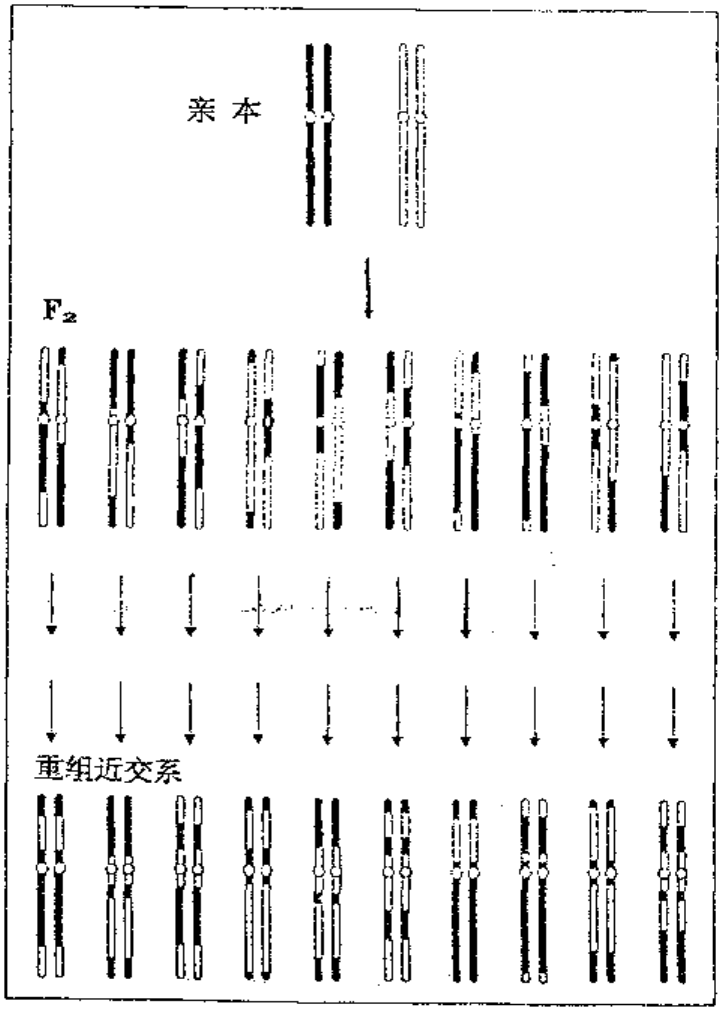
一对基因杂合时，纯合体占到99%，为5对基因杂合时，纯合体占到96%，15对时为89%，30对时为79%，50对时为68%，100对时为46%。

当杂合体中的杂合基因之间存在连锁时，自交同样能使杂合的基因型逐代趋于纯合，只是纯合的速率直接受重组率的影响。重组率愈低，在同一自交后代中纯合体占有的比率愈多，即纯合的速度越快，反之亦然。如果重组率接近于0或完全连锁，自交纯化的速率即接近或等于一对杂合基因自交纯化的速率。如果重组率接近于50%，自交纯化的速率势必接近于两对独立遗传基因自交纯合的速率。据推算，当两对杂合连锁基因的重组率为10%时，自交一代的纯合体占41%，重组率为20%时，纯合体占34%，40%时占26%，45%时占25.26%。

近交所产生的不同RIL品系，构成了分离得以固定的永久性群体，即对于亲本等位基因连锁区段的不同组合方式，每个RIL是固定的(如图5.4)。这种永久性分离群体是建立遗传图谱和进行基因组其它研究的好材料。

对于多基因控制的数量性状，随着基因的逐代纯合，群体中数量性状的均值将由于显性及其互作的消失而逐渐降低到接近亲本平均值的水平。群体中数量性状的方差将随着自

图5.4　RIL的产生



交世代的进展而发生变化，其变化的方向将取决于影响该数量性状的有关基因的效应及其互作。在不同的数量遗传模型下，一粒传群体不同世代均值及方差的变化如图5.5和5.6所示。

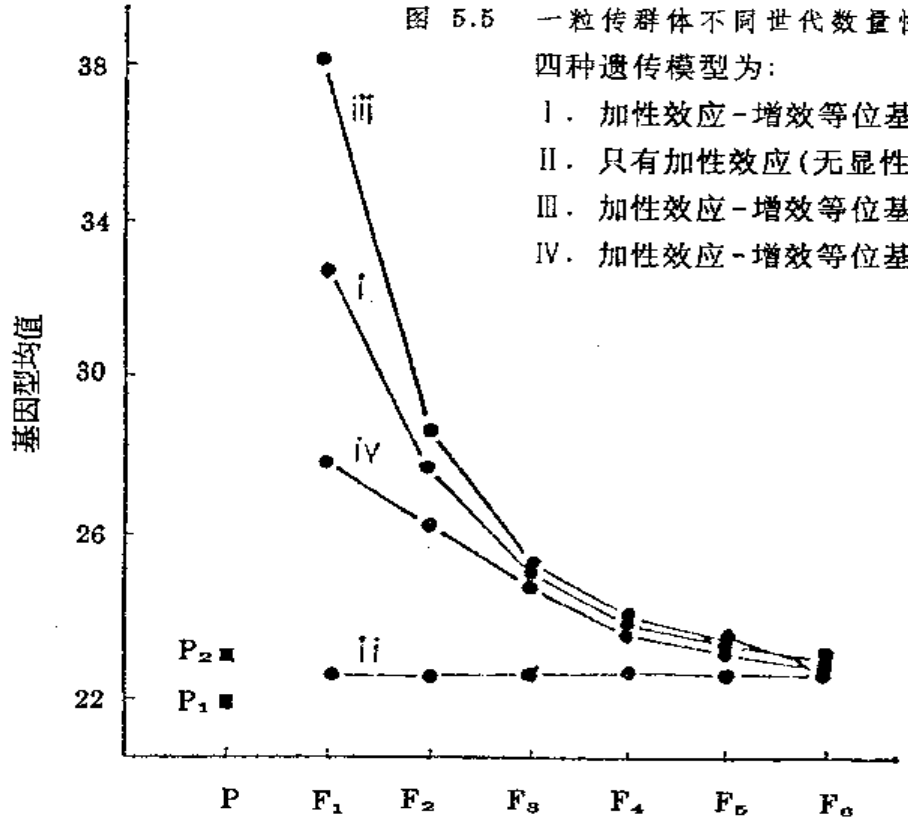


图 5.5 一粒传群体不同世代数量性状均值的变化

四种遗传模型为：

Ⅰ. 加性效应-增效等位基因为完全显性

Ⅱ. 只有加性效应(无显性)

Ⅲ. 加性效应-增效等位基因为完全显性-互补互作

Ⅳ. 加性效应-增效等位基因为完全显性-重复互作

动物中RIL的生产大约需要20代的同胞交配才能达到可以利用的纯合水平，由此可能导致近交系生活力降低，保存十分困难。首先应用于RIL作图的小鼠，其RIL家系的数量一般是较小的。这种较小的家系所造成的问题通常可以利用几个家系的综合资料来克服。与此相反，高等植物达到纯合所需的世代比同胞交配要少一半，但对于有些物种，如烟草，油菜等的一些种，其自交不亲和性可能限制RIL品系的生产。由于植物保存所需费用较小，而且便于保存，因而可以获得比动物大得多的群体。

选择所造成的偏倚，是RIL培育过程中应该注意的另一个问题。C. W. Stuber等在培育玉米RIL的过程中，选取每一近交系植株第一果穗上的种子进行繁殖。但B. Burr等发现由于无意思地选择作用可使RIL中的剩余杂合性为理论值的2倍。任何程度的杂合性都将随着杂合基因型固定为某种亲本等位基因类型而导致后续RIL世代中某种程度的遗传漂变。

RIL是连续近交所产生的。根据物种的繁殖方式的不同以及连续近交程序上的差异，RIL的产生可分为几种不同类型。

1. 全同胞交配

对于异交生物，最极端的近系交配是全同胞交配(full sib-mating)。即同父同母后代间的交配。由于异交生物本身是一个杂合体，因此在产生杂种F_1之前，应对亲本进行多

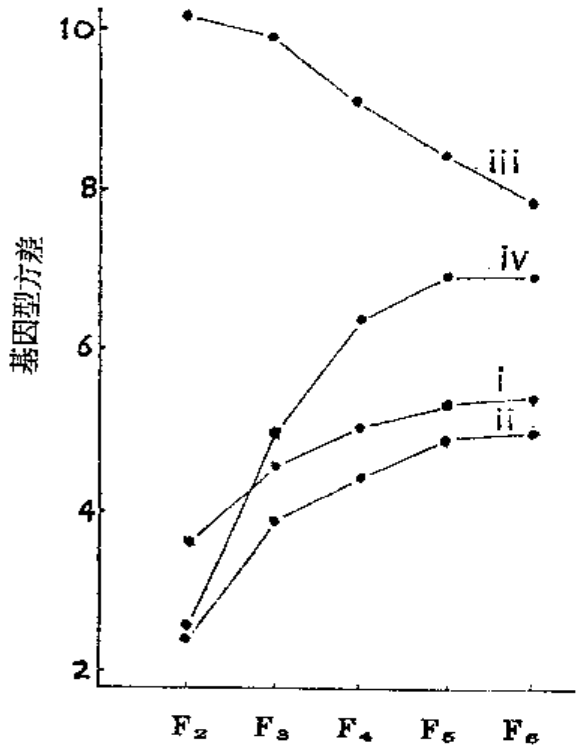


图5.6 一粒传群体不同世代数量性状方差的变化
(四种遗传模型同图5.5)

5.3.2 RIL的产生途径

代近交，产生纯合的近交系，再利用近交系配制杂种，在杂种后代中进行个体间的交配。这种交配方式持续到后代为纯合品系时为止。

2. 自交

对于自花授粉的植物，它们的品种基本上是基因型纯合体，可直接用于配制F_1杂种，然后进行连续自交。根据对杂种后代的处理方式的不同，可将产生RIL的方式分为混合种植法和SSD法。

(1) 混合种植法(Bulk method)

从F_1代开始，将杂种后代混合种植，直到F_5～F_8时，再将各单株的后代种成株系。这种方法可能使杂种群体经受较强的自然选择，从而使某些弱势基因型被淘汰。

(2) 一粒传法(Single seed descent method)

此法由加拿大学者 Goulden于1941年首创。这种方法就是从F_2代开始，从每一植株上选取一粒(或少数几粒)种子进级到下一世代，直到F_5～F_8代绝大部分单株基因纯合时为止，然后收获单株上的全部种子形成SSD株系，组成RIL群体。

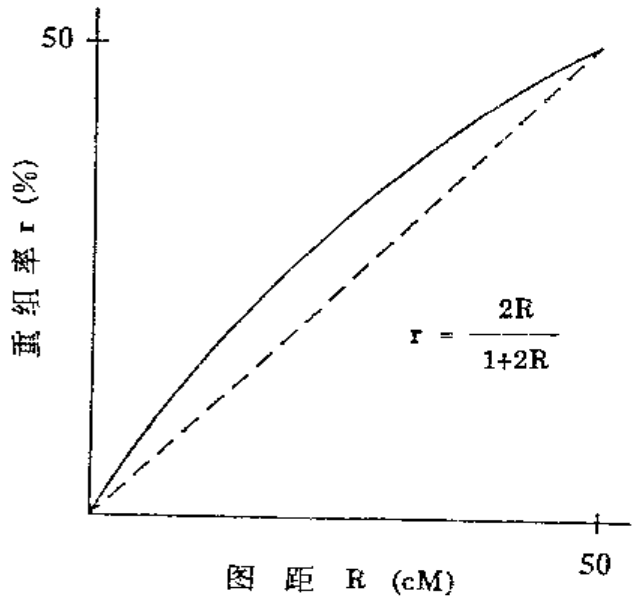
5.3.3 RIL群体中图距与重组率的关系

在理论上，无论自交(近交)多少代，在最终形成的RIL品系中，仍然存在不同基因座位的杂合体。从上述自交的遗传效应分析中，我们已经看到了这类杂合体在群体所占的理论比例。利用RIL进行遗传图谱的构建，所应用的是纯合的RIL品系。与只经历一次减数分裂的F_2和回交一代群体不同，RIL在达到纯合化前已经历了多轮的减数分裂，因而连锁基因

间具有更大的重组机会。这种特性是 Haldane和 Waddington (1931)在近交群体的研究中发现的。对于紧密连锁的标记，RIL群体中所观测到的重组体的数量是一次减数分裂所产生的群体的两倍。在作图的初级阶段，这种扩张性重组使得难以检测到基因之间的连锁。但一旦建立了基本的连锁关系之后，这种扩张性重组使得更容易检测到基因座位之间的不等位，而且可以更精确地估计图距。这是因为图距的置信区间是所观测的重组率的函数，即随着所经历的减数分裂次数的增加，就可以有更多的机会检测到两个连锁标记之间的重组(图5.7)。

图距 R 与重组率 r 之间的关系：在一次减数分裂的群体中，重组率 r 与图距 R(以cM表示)是直线关系（图5.7 中虚线)。对于自交产生的 RIL，当图距很小时，由于 $R=r/(2-2r)$,故 r 接近于2 R（图中实线)。对于同胞交配产生的RIL，达到纯合所花时间更长，这种关系的偏离就更厉害，当图距很小时，r 接近于4R。

图 5.7　RIL群体中图距(R)与重组率(r)之间的关系



5.3.4　利用RIL构建遗传图谱

利用RIL群体所得分子标记数据进行连锁作图是容易的。采用Haldane和Waddington (1931)的公式 $R=r/(2-2r)$可以将 RIL群体的重组率转换为图距。这里没有与一次减数分裂过程中双交换事件相对应的作图函数可以利用。在RIL群体中连锁可以有把握确定的图距范围内，重组率和图距之间存在线性关系(Silver 1985)。基因座位排序的原则是通过排定紧密连锁的标记座位使这些基因座位之间的总图距达到最小。

由于机会的原因，可能将非连锁的基因座位判定为连锁。这种虚假的连锁可以通过下列方式加以判别。一是与一个已定位基因表现虚假连锁的标记不会与同一连锁群上的其它标记表现连锁。二是在一套RIL家系群中所揭示的有疑问的连锁不会在第二套家系中发现。小鼠遗传学家讨论了利用小群体RIL工作时不能检测出真实连锁以及判断虚假连锁的情形。Silver(1985)给出了一个表格，由此可以获得同胞交配产生的 RIL中图距估计值的95%和99%置信限。将这些置信限与同样大小的F_2和回交一代群体的二项分布表进行比较，发现

当重组率较低时，RIL群体的置信限较小。Taylar(1978)计算表明，当 $R \leqslant 0.125M$ 时，同胞交配所得RIL家系比从单一减数分裂所得群体，更能有效地估计图距。采用Taylar的这一方法，可以计算出自交所得的 RIL在 $R \leqslant 0.23M$ 时更能有效地估计图距。

在植物尤其是玉米中目前有大量可用于构建连锁图谱的RIL群体。表5.3给出了其中一部分。Burr等(1988)报道了利用玉米T232×CM37和CO159×TX303所发展的两个RIL群体的RFLP图谱。其数据库含有334个已定位的基因座位，总图距1460cM。其中220个在T232×CM37和CO159×TX303中均存在等位基因的差异。他们比较了两套RIL群体所获得的图距估计值，并将这些估值与公认的图谱进行比较。发现估值的差异可高达2倍，但这些变异仍落在置信区间内，但一部分可能是特定区域内重组率存在真实遗传差异的结果。除畸形染色体10以外，在玉米中不存在已知的染色体重排多态性。因此，在两个RIL群体间没有观察到图距的显著差异是不足为怪的。

表 5.3　玉米中可供利用的RIL群体 (Burr等 1988)

亲本组合	RIL品系数	近交代数
T232×CM37	48	8
CO159×Tx303	41	8
CO159×Tx303	160	5
Mo17×B73	44	7
PA326×ND300	74	5
CK52×A671	162	5
CG16×A671	172	5
CH593-9×CH606-11	101	5
CO220×N28	173	5

采用两个或多个RIL群体比一个群体进行作图更为有益。这是因为：(1)在有些情况下，在一个RIL家系中不易检测到多态性，而在另一些家系中则较容易。(2)在一个RIL家系中所检测到的微弱连锁的证据可以通过检测其它的RIL家系来证实或排除。(3)对于在多个RIL家系均存在分离的标记，其重组资料可以合并起来当作一个大家系处理。因此采用多个RIL家进行作图是有益无害的。

5.4　NIL群体

利用近交方法产生的近等基因系(Near-isogenic lines, NIL)是一系列回交过程的产物。在这里我们先讨论回交的遗传效应及其所产的NIL，然后介绍产生NIL的其它方法。

5.4.1　回交及其遗传效应

回交 (Backcross)是双亲F_1或杂种的其它后代与亲本之一杂交的方式。在育种中，当

某一优良品种缺少一二个优良性状时，常可采用回交的方法将该优良性状从外源品种转移到优良品种中去。用于多次回交的亲本称为轮回亲本(Recurrent Parent)，记为RP，因为它是有利性状(目标性状)的接受者，所以又称为受体亲本。只在第一次杂交时应用的亲本称为非轮回亲本，由于它是目标性状的提供者，故又称为供体亲本(Donor Parent)，记为DP。

回交的结果，将不断提高回交后代中轮回亲本的基因血统，而不断减少供体亲本的基因血统。在一对基因控制目标性状的情况下，在回交第 t 代非轮回亲本的基因在后代中所占比例为 $(1/2)^t$，而轮回亲本所占比例为 $1-(1/2)^t$ (图5.8)。

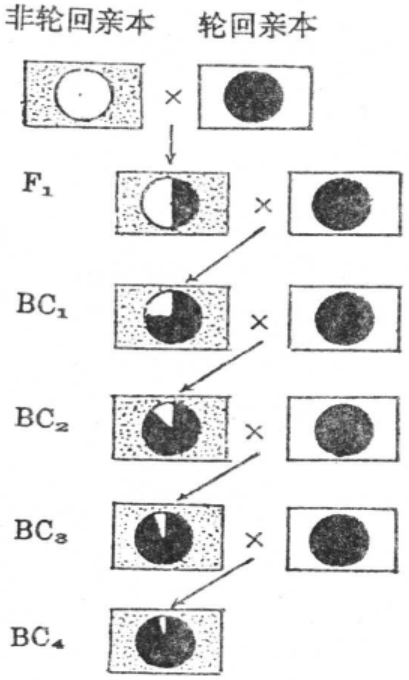


图5.8 回交的遗传效应示意图

在基因不存在连锁的情况下，如果双亲间有 k 对基因的差异，则回交 t 次后，轮回亲本基因的纯合体比例为：

$$[1-(1/2)^t]^k$$

即如果回交5次($t=5$)，当 $k=10$ 时，其纯合体所占比例为72.8%。与自交情况下不同的是，回交的结果使其后向代轮回亲本方向纯合，而自交时，纯合体包括了两种亲本类型及重组类型。

当供体亲本的目标性状基因与其附近的基因存在连锁时，则轮回亲本置换供体亲本基因的进程将要减缓。其减缓程度依连锁的紧密程度而异。在不加选择的情况下，轮回亲本的等位基因置换供体亲本的不良基因获得重组型的机率为 $1-(1-r)^t$。故回交次数越多，重组机率就越高。

当目标性状的基因为显性时，在回交过程中只有那些目标等位基因仍处于分离状态、在表型上与轮回亲本最为相似的个体，才被选择用于与轮回亲本进行回交，其回交过程一

直持续到新培育的目标品系在理论上除了含有目标基因的染色体片段外，与RP几乎等基因时为止。由此得到的回交后代再自交一次即得到回交近交品系(Backcross inbred line, BIL)。BIL又称为RP的近等基因系。

当回交导入的目标性状为隐性时，可将回交一代自交，在分离的自交后代中选株回交。或在回交一代中作较多的回交，同时将回交株自交，将回交与自交后代对应种植，凡是自交后代在目标性状上呈现分离者，说明其相应的回交后代中必有一些个体带有目标性状基因，就可在该后代中继续选株回交并自交；而自交后代不出现分离的，其相应回交后代即被淘汰。

根据回交后代中轮回亲本回复的预期速度$[1-(1/2)^t]^k$，为了保证在回交近交品系(BIL)中消除所有DP基因组，在理论上需要进行无限次的回交。但目前可供利用的BIL都是育种家用不到10代(一般5～6代)回交育成的。因此，任何特定的基因组很可能在其中几个基因座位上含有DP的等位基因，这就是将BIL称为NIL的原因。

5.4.2 NIL作图策略

由于基因连锁的结果，在回交导入目标性状基因的同时，与目标基因连锁的染色体片段将随之进入回交子代中。这种现象称为连锁累赘(Linkagedrag)。NIL作图技术的基本思路是鉴别位于导入的目标基因附近连锁区块内的分子标记，借助于分子标记定位目标基因。通过确定RP、不同NIL衍生物及相应DP三者的标记基因型可以分析分子标记和目标基因之间的连锁。当NIL与DP具有相同的标记基因型，但与RP的标记基因型不同时，就获得了连锁的可能证据(图5.9)。

NIL作图技术的成功应用在于：假定在与目标基因相侧连的DP染色体区域，RP和DP之

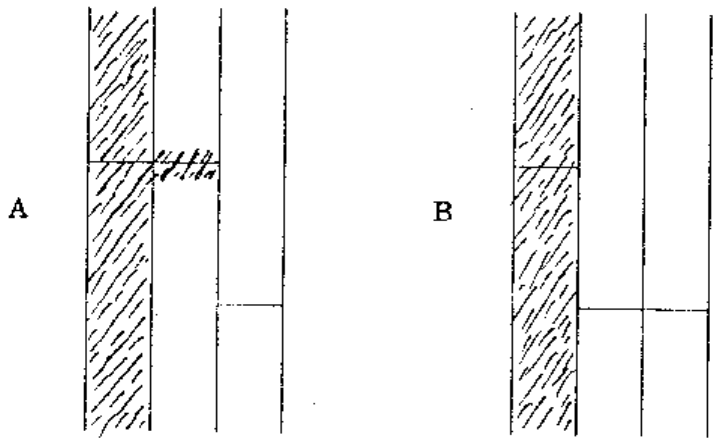


图 5.9 以NIL及其DP和RP为材料通过分子标记定位目标性状基因
A：近等基因系的带与供体亲本相同，可能连锁
B：近等基因系的带与轮回亲本相同，不连锁

间存在分子标记的等位基因差异。显然，检测到这种标记的机会取决于被导入的染色体片段的长度，并且这种机会将随培育NIL中回交次数的增加而降低。而且，这种机会还取决于RP和DP基因组之间分子多样性的程度。当RP和DP分别属于栽培种和野生种时，更有可能发现多态性的分子标记。另一方面RP和DP的亲缘关系越密切，其多态性的分子标记就越少。

利用分子标记，通过筛选大量探针和采用多种限制性酶，可以提高与目标基因连锁的染色体片段内分子多态性的检测机会。在NIL与RP同源的DNA区域，杂交探针将显示同样的带型。相反，在导入片段内的基因组区域，与某种特定限制性酶相结合的杂交探针将在RP和NIL之间揭示出差异。为了寻找这种差异，以5到10个克隆为一组进行同步检测有利于分析大量的探针（Young等1987）。因此，Southern杂交的次数可以大大减少。这些探针可以是随机基因组克隆，最好是单拷贝顺序，但没有必要是过去已经定位的。

将此方法应用于大量的NIL时，这种近等基因系作图技术在检测经典标记与分子标记之间的连锁将是非常有效的。

5.4.3 NIL作图的理论问题

NIL作图技术的一个可能的误差来源就是RP可能不仅在带有目标基因的染色体片段上有差异，而且对于散布于整个基因组的其它DP来源的DNA序列也有差异。这是因为对于有限的回交世代 t，不与导入基因相连锁基因座位上DP等位基因的淘汰率为$(1/2)^t$。显然，这些DP DNA序列可能潜伏着假阳性标记，即实际上不与目标基因连锁的分子标记。Muehlbauer等(1988)的理论计算表明，在所假拟的含有20条染色体，每条长50cM的物种中，随机选择的100个DP来源的分子标记中，有4个预期将滞留在5次回交产生的NIL中。这4个滞留标记中，预期将有1～2个不与目标基因相连锁。这一结果是在以下假定下获得的：在回交过程中，不对RP表型进行选择，即随机选择被转移的标记基因座位处于杂合的子代与RP回交。但在实际育种计划中，育种家总是挑选与RP具有最大相似性的导入标记的杂合体进行回交。对此，我们将对这两种情形分别加以讨论。

1. 没有RP表型选择的回交导入

假定某植物有 n 条染色体，每条长 L 摩(M)，经过 t 次回交将位于某一染色体中部的经典标记(目标基因)从DP转移到RP中。假定100个标记在DP和RP间存在差异，且它们在各染色体上随机分布。则在回交后代中标记染色体(被转移的经典标记所在的染色体)上供体亲本染色体所占比例 U_{Mb} 为：

$$U_{Mb}=2\ (1-e^{-bL/2})/b\ ]/L$$

其方差为

$$V_{Mb}=(2/L^2)\{[2-(bL+2)e^{-bL/2}]/b^2-[(1/b)(1-e^{-bL/2})]^2\}$$

20条染色体（$n=20$），每条长50cM（$L=0.5$），回交5次（$t=5$），$b=t+1=6$，则回交后代标记所在染色体上供体亲本染色体所占比例为：

$$U_{Mb}=0.5179\pm0.2498=51.79\%\pm24.98\%$$

即可以预期标记所在染色体的51.79%是来源于供体亲本的。对于一个50 cM的染色体，这个比例可转换为遗传图谱长度为$0.5179\times50cM=25.90\pm12.49cM$。

回交子代中非标记染色体上供体亲本染色体所占比例 U_{Nb} 为

$$U_{Nb}=(1/2)^b=0.015625=1.56\%$$

其方差为

$$V_{Nb}=(1/2)^b(2/L)[(1/b)(1-e^{-bL/2})]-[(1/2)^b]^2$$

相当于遗传图谱长度为 $0.015625\times 50cM=0.78\pm 4.43cM$。

因此，回交子代中，供体亲本来源的染色体占全部染色体的比例 U_{Tb} 为

$$U_{Tb}=[L\cdot U_{mt}+(n-1)\ L\cdot U_{nt}]/(nL)$$

其相应的方差为

$$V_{Tb}=[(L)(V_{Mb})+(n-1)(L)(V_{Nb})]/[(nL)^2]$$

因而

$$U_{Tb}=(0.5179\times 50+(20-1)\times 50\times 0.015625)/(20\times 50)$$
$$=0.04074\pm 0.1028$$

相当于遗传图谱长度为 $0.04070\times 20\times 50=40.74\pm 1.06cM$。这40.74cM的DP基因组中，有25.9cM(占64%) 位于导入标记所在染色体上的连锁区段内，而 40.74-25.90 =14.84 cM(36%) 位于非导入标记所在的染色体上。因此，在 BC_6 衍生的NIL中，100个随机选择的标记中大约含有 $100\times 0.04074=4.074$ 即4个DP的标记等位基因。

这4个标记中约有 $4.074\times 25.90/40.74=2.592$ 即2个或3个位于导入标记所在的连锁区段内，其余的 4.074-2.59=1.484 个则位于非导入标记所在的染色体上而表现为假阳性。

在理论上，100个随机选择的分子标记座位中，将有5个基因座位处于构成总基因组二十分之一的任何特定染色体上。很显然，与采用随机挑选分子标记进行分析的其它方法相比，NIL作图技术所需分析的分子标记数大大降低。对于具有20条染色体的物种，BC_4 (或更高)世代所衍生的NIL中所滞留的标记有一半或更大部分将与导入标记相连锁(表5.4)。同样大小 (10 M)的基因组，由数目更多或更少的染色体组成时，滞留在NIL中的连锁和非连锁DP基因组所占的比例也会发生变化。总染色体越少，标记染色体在总基因组中所占比例就越大，因此连锁与非连锁DP基因组之比（或连锁与非连锁DP基因座位之比）就越高(表5.4)。

发生“假阳性”标记的概率随着回交次数的增加而减少，因为NIL中导入标记所在的染色体片段长度的降低速度低于与导入的标记基因不连锁的DP基因组。提高该方法效率的另一手段是比较从同一对RP和DP独立产生的两个或多个NIL。由于在每一NIL获得同一“假阳性”标记的机会是相互独立的，假阳性出现的概率对于两个NIL时为 p^2，而不是一个NIL时的 p 了。

由于只需要分析少数的基因型(RP，NIL)，NIL作图技术为将一大套随机选择的标记削减到只涉及少数几个标记(其中一些与NIL中的导入基因紧密连锁)提供了适当的方法。对于这少数几个标记，必须利用分离群体和传统的分离分析来证实是否存在真实的连锁，并获得有关基因排列顺序和相间图距的信息。因此，NIL作图技术是对将单个目标基因与分子标记结合为同一遗传图谱的方法的补充而不是替代。

表 5.4　回交产生的NIL中来自供体亲本标记染色体(M)、非标记染色体(N)及全部染色体(T)所占比例及遗传图距 [b=回交次数+1; n =染色体数; L=单个染色体长度(cM)]

参 数			标记染色体		非标记染色体			全部染色体	
b	n	L	U_{Mb}	$L\,U_{Mb}$	U_{Nb}	$L\,U_{Nb}$	$(n-1)L\,U_{Nb}$	U_{Tb}	$nL\,U_{Tb}$
5	1	10.00	0.04000	0.40000	-	-	-	0.04000	0.40000
5	2	5.00	0.08000	0.40000	0.03125	0.15625	0.15625	0.05562	0.55625
5	4	2.50	0.15969	0.39923	0.03125	0.07813	0.23438	0.06336	0.63360
5	8	1.25	0.30594	0.38243	0.03125	0.03096	0.27344	0.06559	0.65586
5	10	1.00	0.36717	0.36717	0.03125	0.03125	0.28125	0.06484	0.64842
5	20	0.50	0.57080	0.28540	0.03125	0.01563	0.29688	0.05828	0.58277
5	40	0.25	0.74358	0.18590	0.03125	0.00781	0.30469	0.04906	0.49058
6	1	10.00	0.03333	0.33333	-	-	-	0.03333	0.33333
6	2	5.00	0.06667	0.33333	0.01563	0.07813	0.07813	0.04115	0.41146
6	4	2.50	0.13326	0.33315	0.01563	0.03906	0.11719	0.04503	0.45034
6	8	1.25	0.26040	0.32549	0.01563	0.01953	0.13672	0.04622	0.46221
6	10	1.00	0.31674	0.31674	0.01563	0.01563	0.14063	0.04574	0.45736
6	20	0.50	0.51791	0.25896	0.01563	0.00781	0.14844	0.04074	0.40739
6	40	0.25	0.70351	0.17588	0.01563	0.00391	0.15234	0.03282	0.32822
7	1	10.00	0.02857	0.28571	-	-	-	0.02857	0.28571
7	2	5.00	0.05174	0.28571	0.00781	0.03906	0.03906	0.03248	0.32478
7	4	2.50	0.11427	0.28567	0.00781	0.01953	0.05859	0.03443	0.34426
7	8	1.25	0.22569	0.28212	0.00781	0.00977	0.06836	0.03505	0.35048
7	10	1.00	0.27709	0.27709	0.00781	0.00781	0.07031	0.03474	0.34740
7	20	0.50	0.47213	0.23606	0.00781	0.00391	0.07422	0.03103	0.31028
7	40	0.25	0.66644	0.16661	0.00781	0.00195	0.07617	0.02428	0.24478

2. 具有RP表型选择的回交导入

上述有关公式是在假定回交时对轮回亲本的表型不加选择，随机从回交后代中选择被导入标记座位的杂合体进行回交的前提下得到的。在实践中，为了尽快获得所期望的回交后代，总是挑选与轮回亲本最为相似的标记杂合体进行回交。这种选择回交只有在下列条件下才是有效的：(1) RP和DP在较多的经典标记座位上而不是只在被转移的基因座位上存在等位基因的差异；(2)这些有差异的等位基因控制着回交子代的表型差异(即具有较高的遗传率)；(3) 每一回交子代产生了足够数量的个体，从而为不同表型差异的个体提供

了相当的机会。 如果上述三个条件得到满足，至少是部分得到满足，则RP表型选择不仅可以直接提高被选择的标记座位上轮回亲本等位基因的回复率，而且还可以提高与选择基因座位相连锁的RP染色体片段的回复率。

RP表型选择对前述计算结果有两个重要影响。每一轮回交中DP基因组被淘汰的程度增加，在回交结束时滞留在回交后代中的DP基因组的平均比例减少。由于DP基因组的淘汰程度是回交次数的近似函数，RP表型选择对这两个参数的绝对影响以早期回交世代最大，而在较高的回交世代，这种影响明显降低。

要为RP表型选择推导出在不加选择时得到的那些公式是困难的，因为这些式子必须能够解释影响RP表型选择效应的三个因素(条件)，而这些因素对于产生NIL的特定组合DP×RP都是不同的。而且需要了解与所要选择的RP表型有关的基因是否位于标记染色体上。如果是，它们相对于被导入的标记座位的连锁距离是多少。 显然要了解所有这些是困难的。但是我们可以用两个假设例子来说明RP表型选择对DP基因组滞留于标记染色体和非标记染色体的影响。

如果所有可供选择的基因座位碰巧都位于非导入标记所在的染色体上，则成功的RP选择将使染色体上DP基因组所占的比重大大低于表5.4所列的数据。这一点可以通过表5.5的数据来说明，表中假定遗传率为100%的可选择基因座位处于19条非标记染色体的中部，并且在第一次回交中在所有19个基因座位上成功地完成了RP等位基因的回复。根据这些假定，BC_n衍生的NIL中非标记染色体中滞留的DP基因组的量从14.84cM 减少到3.76cM (表5.5)。

相反，如果可供选择的标记碰巧位于标记染色体，则在基因座位对RP等位基因的成功

表 5.5 BC_n衍生的NIL中来自供体亲本的非标记染色体所占的比例和遗传图距
(t=回交次数； $b=t+1$； 染色体数 $n=20$；每条染色体长0.5cM)

t	b	不作RP表型选择			RP表型选择		
		U_{Nb}	$L\,U_{Nb}$	$19\ L\,U_{Nb}$	U_{Nb}	LU_{Nb}	$19\ L\,U_{Nb}$
0	1	0.50000	0.25000	4.75000	0.50000	0.25000	4.75000
1	2	0.25000	0.12500	2.37500	0.06333	0.03166	0.60159
2	3	0.12500	0.06250	1.18750	0.03166	0.01583	0.30079
3	4	0.06250	0.03125	0.59375	0.01583	0.00792	0.15040
4	5	0.03125	0.01563	0.29685	0.00792	0.00396	0.07520
5	6	0.01563	0.00781	0.14844	0.00396	0.00198	0.03760
6	7	0.00781	0.00391	0.07422	0.00198	0.00099	0.01880
7	8	0.00391	0.00195	0.03711	0.00099	0.00049	0.00940
8	9	0.00195	0.00098	0.01855	0.00049	0.00025	0.00470
9	10	0.00098	0.00049	0.00928	0.00025	0.00012	0.00235

选择，结合对导入DP标记的选择，必然导致这两个基因座位重组事件的发生。假定两个基因座位的遗传率为100%，则预期重组事件将发生在这两个基因座位之间。例如，当两个基因座位的连锁距离为10cM时，则平均而言，重组事件将发生在与两个基因座位等距离的5cM处。事实上，这将把DP染色体片段在导入标记附近的滞留限制在5cM的范围内，而不是在不加RP选择时所预期的12.95cM（即25.90cM的一半）。显然，对于标记和非标记染色体，成功的RP表型选择可以将DP基因组的滞留量减少到表5.4所列的值以下。但是，这种选择对非标记染色体的效应要更大一些。可以预期，位于标记染色体上的可选择基因座位所占的比例将是标记和非标记染色体长度之比（即在20个染色体的物种中可选择基因座位的5%）的函数。其次，对于导入的DP标记的选择是100%有效的，而RP表型选择涉及到遗传率较低的数量性状。事实上，有效选择强度上的这种差异将使重组事件（如果完全发生的话）偏离连锁距离的中点（即远离导入基因座位）。从这两方面考虑，RP表型选择在很多情况下将不至于降低或许会提高NIL作图的效率。

前述计算表明，如果在假设的物种中，由BC_6所衍生的NIL在100个随机选择的分子标记座位上存在RP/DP等位基因的差异，则可以预期在其中96个基因座位上，RP/NIL的等位基因是相同的，而NIL/DP是不同的。而在另外四个基因座位上的情形与之相反。由此可以推断，96个基因座位可能与NIL中被导入的经典标记座位是独立的，而另外四个则与之连锁。

由于回交的随机性，可能有些本来与导入标记连锁的分子标记等位基因被淘汰了，而本来与之独立的分子标记却被滞留下来。在NIL作图中出现这类错误的概率可作如下评价。如果在BC_6的自交一代所衍生的NIL中未观测到特定的DP分子标记，则该标记可能不存在于标记染色体上。这一假定为错误的概率（p是预期在标记染色体上将要回复的RP基因组与所有染色体上将要回复的RP基因组之比的函数），即

$$p = 100\ [\ L(1-U_{Mb})]\ /\ [(nL)(1-U_{Mb})]$$

对于所假设的物种，$p = 100(24.10/959.26) = 2.5\%$。如果对于某一特定的导入标记，有两个独立衍生的NIL可以利用，则在上述情况下发生错误的概率为 $p^2 = 0.0625\%$。

另一方面，如果在NIL中观测到了特定DP分子标记的存在，则该标记可能位于标记染色体上。这种假定发生错误的概率（p是预期在非标记染色体上将要滞留的DP基因组与所有染色体上将要滞留的DP基因组之比的函数），即

$$p = 100\ [\ (n-1)L(U_{Nb})]/\ [(nL)(U_{Tb})]$$

对于所假设的物种，$p = 100\ (14.84)/40.74 = 36.4\%$。这一概率将随回交次数的增加或/和染色体数目的减少而明显降低（表5.4）。如果两个独立衍生的NIL中都带有导入的分子标记等位基因，则假定发生错误的概率为 $p^2 = 13.3\%$，因此NIL作图的效率将大大提高。

5.4.4 应用

有几篇研究报告已成功地应用了NIL作图技术（Young等1988，Gupta等1989）。Young等在番茄中鉴别了与控制烟草花叶病毒抗性的基因相距不到0.5cM的分子标记。这一实验结果表明NIL作图技术对于寻找与目标基因紧密连锁的分子标记是十分有效的。这类紧密

连锁的标记适合于标记辅助的性状选择，而且对于染色体步行克隆目的基因也是绝对必需的。

番茄抗烟草花叶病毒的基因 *Tm-2*，通过连续回交的方法转移到了不同的栽培品种中，从而得到了一系列轮回亲本不同的NIL。以这些NIL为材料，Young和Tanksley (1986)研究了 *Tm-2* 两侧导入片段的大小。发现不同NIL 材料中，导入片段的长度从最短的4cM 到最长的51cM（包含了第九染色体的整个短臂）。结果表明，由于在回交过程中很少对抗性基因附近区域的理想重组体进行选择，导致回交育种方法不能有效地减少抗性基因附近的连锁片段的大小。他们建议通过检测目标基因与其附近分子标记的重组来快速有效地减少与导入有关的连锁累赘程度。采用这一方法，在两次回交中所获得的导入片段与不加选择下100代回交所得导入片段一样小。

Young和Tanksley(1989b)还根据Hanson(1959)和Stam和Zeven(1981)的理论，比较了不同回交代数下，供体亲本中与目标基因相连锁的以及不连锁的DNA片段的大小。结果表明，对于10条染色体，每条长100cM 的基因组，在前4次回交中，来自供体亲本的非连锁DNA 的比例高于其余的连锁DNA 的比例。但4次回交后，由于连锁累赘所带来的供体亲本DNA 的比例大大超过来自非连锁区域的。在第10 次回交世代中，连锁的供体 DNA 预期将超过非连锁的供体 DNA 50 倍。这表明了连锁累赘对回交育种的影响。

Paran等(1991)将NIL用于鉴别与莴苣霜霉病抗性基因 *Dm* 相连锁的RFLP及RAPD标记。采用了两对在 *Dm1* 和 *Dm3* 上有差异，一对在 *Dm11* 上有差异的NIL为材料，用500个 cDNA 和212 个随机寡核苷引物对NIL 进行多态性检测。结果发现4个RFLP、4个RAPD标记与 *Dm1* 和 *Dm3* 区域连锁，6个RAPD标记与 *Dm11* 连锁。即有1% 的DNA克隆，不到1% 的PCR 扩增产物在所筛选的 *Dm* 区域不同NIL 间是多态性的。这低于3% 的最大预期值。该最大预期值是在以下假定下计算的：

(1) 在导入区域NIL之间存在100% 的多态性

(2) 回交时不作轮回亲本表型选择

(3) 所筛选的标记在整个基因组上随机分布

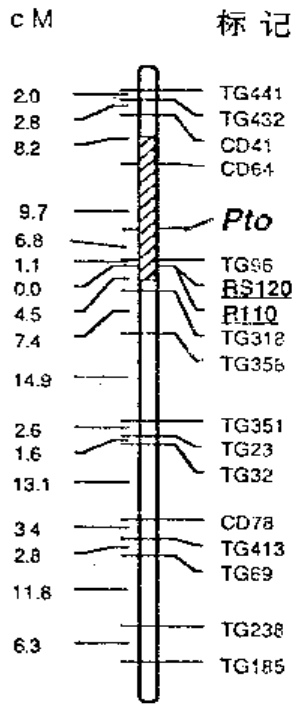
所观测到的低水平多态性反映了上述假定在某种程度上未能满足。

利用同一染色体区域的一个以上的NIL，可以减少在未连锁区域检测到标记的机会，增加在导入片段中检测到多态性的概率。在NIL间表现多态性但不与目标基因连锁的标记，将导致假阳性。这种假阳性的产生是由于非连锁供体染色体片段在回交过程中随机导入的结果。 这些染色体片段可能通过增加回交次数或借助于标记的选择来缩小(Tanksley等1989)。

Martin等(1991)采用RAPD 方法与 NIL 技术相结合，快速鉴定了与番茄茎腐病抗性基因 *Pto* 相连锁的 DNA序列。利用144个随机引物对第5染色体上 *Pto* 基因有差异的一对NIL进行筛选，7个扩增产物存在于一个NIL中而在另一个NIL 缺乏。对其中4个扩增产物作进一步分析，有3个被分离分析证实与 *Pto* 基因连锁(图5.10)。

在NIL中目标基因特定区域内获得一个有差异的标记的概率不仅取决于RAPD分析中所用引物的数目，还取决于以图距单位表示的基因组的大小以及目标基因附近NIL之间DNA

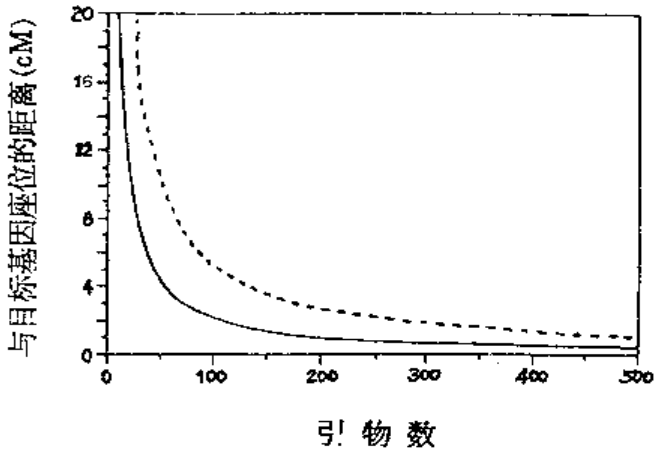
图5.10　番茄第5染色体的RFLP图谱
（示 *Pto* 基因及其连锁标记）



利用NIL技术发现了7个PCR扩增产物的差异，其中3个（转换为RFLP标记R110、R120和TG96）被分离分析证实与 *Pto* 连锁。阴影区域为来自供体亲本的染色体片段，空白区域为轮回亲本的染色体，与TG96相侧连的RFLP标记的分析提供了有关 *Pto* 基因附近导入片段大小的信息

序列多样性的程度。如果NIL间序列非常相同，则一个引物是不太可能检测到PCR多态性的，即使引物扩增发生在目标基因附近也是如此。图5.11中假定NIL中目标基因附近的扩增100%产生可检测的PCR多态性，特定数目的引物可能获得的期望图距及95%的置信图距将随可检测多态性比例的减少而成比例增加。基因组大小的增加也将导致所估计的图距成比例增加。

图5.11　目标基因与PCR标记之间的最小图距和RAPD分析中所用引物数的关系



实线为期望的最小距离=g/[2(nx+1)]，其中g为基因组大小（番茄为1500cM），n=引物数，x为每个引物的平均PCR产物数（这里为4）。95%置信限下的图距（虚线）=(c/2)(1-0.05$^{1/n}$)。

5.4.5　产生 NIL的其它方法

除了可通过连续回交获得 NIL外，亦可通过连续自交产生 NIL。其基本方法是利用

带有目标性状的品系与另一遗传背景优异的其它品系杂交，在自交后代中对目标性状和其它性状进行选择，并注意在选择过程中始终使目标性状处于杂合状态。当自交到较高世代后，对目标性状处于杂合状态的个体再进行一次自交，由此分离出目标性状纯合的和对应的不带有目标性状的两类个体(组成 NIL)。

通过突变或人工诱变的途径可能获得与原始亲本只在突变性状上存在差异的一系列突变品系，突变品系与对应的原始品系就构成了一对 NIL。凡是可产生单基因突变或单一性状的多基因突变的方法，都可以用来获得 NIL。不过，这类方法一般难以根据需要定向地产生具有目标性状差异的成对NIL。为了今后研究的需要，在诱变育种和组织培养等工作中应注重突变体的分析和保存。

5.5 几类群体的连锁作图及其效率

分子标记在育种中的应用大多需要对标记和目标基因座位进行精细的连锁作图。对于寡基因性状，其遗传模式可由经典的孟德尔分析来确定，其遗传图谱可以通过分离分析或/和近等基因系的比较来确定。对于一些物种，可以利用非整倍体(Helentjaris等1986a)、染色体代换系或易位材料(Weber和Helentjaris 1989)来简化遗传作图。这些方法主要用于证实分离分析的结果，将连锁群与染色体或染色体臂挂勾以及把分子图谱与靠细胞学、形态学和同工酶资料构建的经典遗传图谱更密切地联系起来。

植物中以分子标记建立的遗传图谱通常是用两个纯合亲本间F_1所产生的分离群体进行的。在大多数情况下，采用的是F_2或一次回交的后代，以单株作为分离的单位。此外，在玉米中采用一粒传产生的随机近交系RIL已构建了RFLP连锁群 (Burr等 1988)。在大麦和水稻中采用双单倍体方法产生的随机近交系为作图群体构建了分子连锁图谱(Heun等1991和Zhu等1993)。

表5.6总结了以各类群体为材料采用最大似然法估计重组率的估计式。除了上述提到的群体外，还考虑了由全同胞交配产生的随机近交系RILf及F_2群体随机交配产生的F_{2S}。并作以下假定：

(1) 各基因座位的分离符合孟德尔遗传下共显性基因的分离比例

(2) 雌雄两性中重组率是相同的，除了F_2和F_{2S}以外，最大似然方程的显式解是存在的。

对于假定(1)的偏离可以由减数分裂的不规性，或配子与合子的选择所引起。在黑麦中报道了与不亲和性基因座位连锁的同工酶标记的偏离 (Wricke和Wehling 1985，Gertz和 Wricke1989)。Heun和Gregorius (1987)推导了不同遗传模式下单基因异常分离的F_2群体重组率的最大似然估计式。

表5.6还提供了由单个分离单位所贡献的估计重组率r的平均信息量 $I(r)$。$I(r)$取决于r值的大小及作图群体的类型。r 的最大似然估值的方差为：

$$\sigma^2_r = 1/[nI(r)]$$

其中 n 为所分析的分离单位的数目， 即群体容量。对于独立分离的基因座位 ($r=0.5$)时， σ^2_r最大；而当$r \to 0$时， σ^2_r收敛于0。

表5.6　不同作图群体两个基因座位间重组率(r)的最大似然估计式及平均信息量$I(r)$的计算式

群体	r估计式	信息量$I(r)$
F_2	r_1：解方程 $\frac{2x}{r}+\frac{2y}{r-1}+(n-x-y-z)[\frac{1-2r}{r(1-r)}]+z[\frac{2(2r-1)}{1-2r+2r^2}]=0$	$I_1(r)=\frac{2(1-3+3r^2)}{r(1-r)(1-2r+2r^2)}$
F_{2S}	$r_2=(3/4)-[(9/16)-r_1]^{1/2}$	$I_2(r)=I1(r^*)\times(9-16r^*)/4$
DH;BC	$r_3=y/n$；$r_4=w/n$	$I_3(r)=I_4(r)=1/[r(1-r)]$
RIL	$r_5=y/(2x)$	$I_5(r)=2/r(1+2r)^2$
RILf	$r_6=y/(4x-2y)$	$I_6(r)=4/r(1+2r)(1+6r)^2$

x:=[$AABB,aabb$], y:=[$AAbb,aaBB$], z:=[AB/ab, Ab/aB], w:=[$Aabb$, $aaBb$] 分别为基因型 $AABB$、$aabb$ 等的观测频率；假定亲本基因型为 $AABB$ 和 $aabb$，以 $aabb$ 作回交亲本；n 为总分离单位数(群体容量)；$r^*=r(1.5-r)$

根据 Allard(1956)，将可以完全区分为10 种可能的基因型的F_2用作比较不同类型群体的标准。该F_2的平均信息量为

$$I^*(r)=r(1-r)/2$$

图5.12给出了相对于 F_2的 $I(r)$下，6种类型群体的相对信息量 $I(r)$。DH和BC群体所能提供的信息正好是完全可分类F_2的一半，因为这两类群体只有单方配子中的重组事件的信息。不能区分两种类型双杂合体的F_2，其相对信息从 $r=0$时的100%几乎直线下降为 $r=0.5$时的50%。

对于紧密连锁的基因座位，RIL提供了几乎与F_2群体相同的信息量。但当 $r>0.15$时，它所能提供的信息量甚至少于DH群体。对于极紧密连锁的基因座($r=0.05$)，RIL所提供的信息量约为F_2群体的2倍。但对于较大的 r 值，其信息量迅速下降，低于所有其它作图群体。相反，对于非常紧密的连锁，F_{2S}比F_2提供的信息多50%，但当 $r>0.17$后，其效率低于不能区分两种类型的杂合体的F_2。由此可见，由于互交、同胞交配或自交所导致的重组机会的增加，当 r 值较小时使信息量增加，而当 r 值居中或较大时，将降低 r 估计的效率。

除了所研究物种繁殖生物学方面的限制因素外，最佳作图群体的选择主要取决于研究的目标。对于具有相当饱和连锁图谱的物种，与标记基因紧密连锁的基因定位可以利用F_2和F_{2S}来有效地进行。尽管RILf甚至可以提高极紧密连锁基因座位的作图效率，但其利用仍受到限制，因为为了获得满意的近交程度必须进行多代近交。

对于第一次构建遗传标记图谱，F_2 群体优于RIL或DH 系，因为从生物统计学的观点看，对于特定的样本n，在0～0.5整个区间，F_2的最大似然估计值r具有一致最小方差。但在实际应用上，如前所述，RIL和DH与分离群体相比具有多方面的优点。

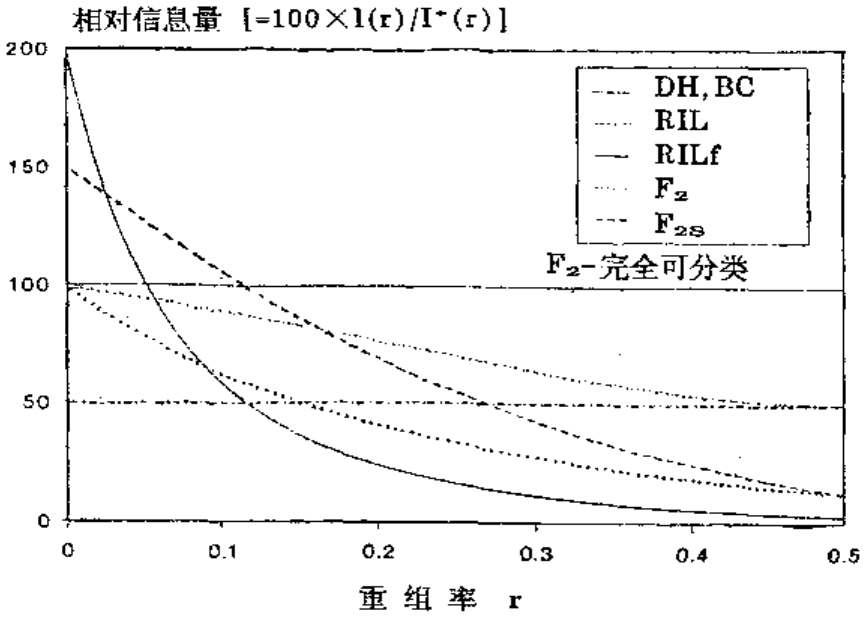


图5.12 不同作图群体每一作图单位所贡献的平均信息量$I(r)$与重组率之间的关系
($I(r)$以完全可分类F_2群体的百分比表示)

(1) 该群体代表着一个永久性的群体，因为所有的等位基因都是固定的，可以无限地用于新标记的作图，并可在研究小组之间共享。

(2) 该群体可以重复进行检验，特别适合于抗性分析，数量性状的分析等。

(3) 其分子标记基因型可以从自交产生的10～20个植株的混合样品的分析（而不是从单株F_2或F_{2s}进行分析）加以准确的测定。

5.6 群体处理方法

为了有效利用遗传群体所提供的信息及最大限度的降低试验分析的成本，现已发展了多种群体处理和分析的方法。这里主要介绍选择性基因型分析和分群分析方法。

5.6.1 选择性基因型分析

在已有的标记-QTL连锁研究中，通常利用一个群体同时进行多个性状的连锁分析。但在未来的研究中将主要针对单个性状。 在这种情况下，如果只对抽样群体中高表型和低表型的个体进行有选择的基因型分析(图5.13)，就可以在保持特定的统计学效率的前提下大大减少分析的个体数(Lebowitz等1987)。考虑到标记与性状连锁所造成的结果，当分析未知分布的性状或者质量性状时，根据分布之尾标记基因频率的变化进行连锁分析是恰当的。Lander和Botstein (1989) 证实对于连续变异的性状，当根据群体中两尾端的个体的数量性状进行分析时，其效率将显著提高，他们将最大似然法应用于数据的分析，并将这

类方法称为“选择性基因型分析 (Selective genotyping)”。

Darvasi和Soller (1992)的研究表明，从使分析成本达到最低考虑，选择性分析中的最佳选择比例取决于对个体进行所有标记分析所需成本与进行表型分析所需成本之比。但是当研究单一性状时，只需选择占群体总数25%的两尾个体进行分析，对更多的个体进行分析是毫无意义的。在选择群体中所观测到的标记基因型之间数量性状的差异将大大高于对整个群体分析时该QTL的实际基因效应。他们还提供了一个公式和曲线图，以便将选择性基因型分析时所观测到的差异转换为实际的基因效应。

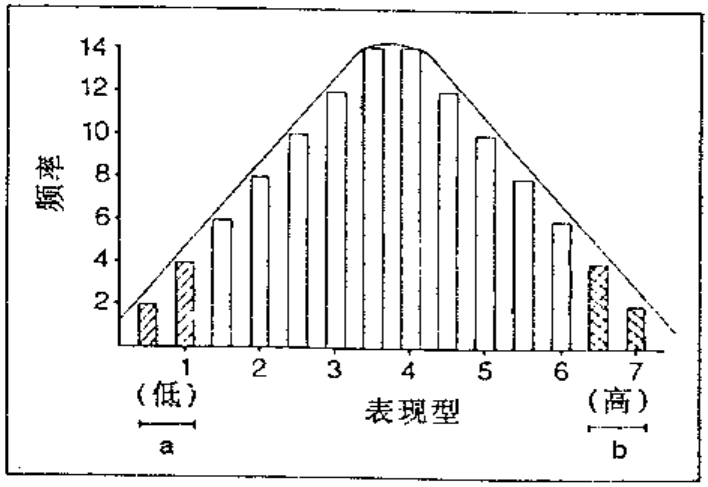


图5.13 选择性基因型分析示意图(阴隐部分的个体被选择用于分析)

Mansur 等(1993)将选择性基因型分析和分群分析的方法应用于QTL的基因定位。他们以两个大豆品种所衍生的284个RIL为材料，从中选择具有极端表型的20个高表型系和20个低表型系，构成两个DNA池 (Pool)，对与亲本RFLP等位基因的片段相结合的放射性探针进行定量分析。当某一RFLP标记与QTL连锁时，来自特定表型的DNA池中某一亲本的等位基因将占优势，而另一亲本的等位基因将与相反的表型相关联。当不存在连锁时，对于特定的性状，放射性强度与两种等位基因的关联(或者对于特定的等位基因，与两类表型的关联）是同等的。

5.6.2 分离群体分群分析方法

Michelmore等(1991)发展了一种分离体混合分析或分群分析法(Bulked segregant analysis)，以用于快速鉴别与特定基因或染色体区域相连锁的标记。基本原理如下：从某一杂交组合的分离群体中产生两个混合的DNA 样品。每一样品池或混合物由特定性状或基因组区域相同而其它非连锁区域完全随机的同类个体所组成。两个混合样品具有完全相对的性状（如一个表现抗病而另一个表现感病）。因此，这两个混合物在所选择的区域在遗传上是相异的，而在其它区域则可能是杂合的。可以针对不同的基因组区域或利用不同的分离群体来组成这类混合体。然后利用 RFLP或 RAPD 分析在两个混合体间寻找有差异

的分子标记。两类混合体间表现出多态性的分子标记在遗传上可能与构建该混合体所用性状的基因座位相连锁。在一般情况下，这种可能的连锁还要通过分离群体的分析加以验证。

Michelmore等 (1991)利用分离体混合分析的方法鉴别出与莴苣霜霉病抗性基因相连锁的3个RAPD标记。与等基因系分析相比，分离个体混合分析在鉴别基因组特定区域的标记方面具有几个优点。混合体可以从现存的分离群体构成，而不必进行等基因系的培育；通过混合分析所鉴别的多态性在形成混合体的群体中总是处于分离状态并能进行定位；利用在前一轮分析中所鉴别的基因座位上有差异的新的成对混合体进行分析，这样可能通过多轮混合分析来实现遗传步行(Genetic walking)。这一方法对于自交和异交物种均有广泛的应用前景。

以上是根据单个标记对群体进行分群混合的分析方法。Giovannoni等 (1991)提出了一种根据两个或多个连锁的RFLP标记所包含的染色体区间上两种不同的等位基因纯合体进行分群混合的方法。对分离体进行分群后，利用随机寡核苷酸引物和PCR技术在两个混合物之间进行多态性筛选。多态性的DNA谱带就代表了与被检测区间相邻或区间内的DNA序列。这样可以快速分离出由RFLP标记所确定的各个基因组区间上的特定分子标记。具体作法是，从分离群体中选择已知标记区间内纯合的两类个体，一类为一种亲本的纯合体，另一类为另一亲本的纯合体。由两种纯合体组成两种DNA的混合体。这样的两个DNA混合体在目标区域内或与目标区域相邻的所有基因座位上都是纯合的，但两个DNA混合体之间在纯合的目标区域其亲本来源是不同的。这样就为在目标区域上检测多态性的分子标记提供了基础。当每一混合体所含个体的数量足够大时，就可以有较高的概率含有同时来自两个亲本的基因组序列，这一结果对于除了连锁非常紧密且位于目标区域之内的那些基因座位外，所有其它的基因座位都是成立的。然后用两种混合DNA分别作为随机引物扩增的模板。利用电泳分析比较来自每一混合DNA的扩增产物，以寻找两种混合DNA之间的多态性。由于除了目标区域内或目标区附近外的其他基因组序列在两种混合体之间基本上是同质的，因此只有当引物引发于目标区间内或区间附近时才会产生多态性。最后对多态性的PCR产物(或采用纯化的多态性DNA带作分子杂交所检测到的RFLP)进行分离分析，进一步证实标记在目标区域的位置。Giovannoni等(1991)以番茄第10染色体上一个15cM的区间和第11染色体上一个6.5cM的区间为例对这一方法进行了验证。这两个区间上存在着控制番茄落果和成熟性的基因。每一区间由7～14个F_2个体组成混合DNA，用200个随机引物进行筛选。结果发现了三个多态性的标记，其中两个被证明与所选择的区间是紧密连锁的。

与利用单个标记进行的分群混合法相比，选择和利用遗传区间近等的混合DNA具有4个优点。1. 对于构成近等混合DNA的群体中不存在多态性的基因座位，仍然可利用其相邻的基因座位进行分析；2. 对于着丝粒等无法在分离群体中直接分析的染色体结构和功能性基因座位，只要它们所在的区间已经被确定就仍然可以进行分析；3. 与单个标记相比，选择区间两边的标记可以更有效地填充遗传图谱中的空缺。同时选择目标区间内两个或更多的RFLP标记处于纯合状态而与区间相邻的RFLP标记处于杂合状态的个体组成混合体，可以增加多态性扩增产物来自区间内序列的可能性。4.来自同一分离群体的DNA样本可以采用不同方式混合，从而对基因组上的任意区间进行分析。

决定混合体中所含个体数多少(亚群体大小)的因素包括:(1) 确定目标区间的两个连

锁标记间的距离；(2) 所要求的统计效率。Giovannoni等(1991)讨论了最佳的区间和混合个体数。利用更多的个体组成混合体将使两个混合体间除了目标区域内的或与目标区域紧密连锁的基因座位外等位基因没有差异的可能性。但是，随着混合体所含个体数的增加，在混合体中个体在区间内发生双交换的概率也将增加。如果在一种混合体中的确有的个体发生了双交换，这将减少或消除在双交换发生的区域找到PCR标记的机会。当区间较大且只根据与区间相邻的遗传标记进行选择时，这就显得特别重要。如果能够提供在该区间上遗传标记的其它信息，则可以减少双交换个体的概率。

这一方法除了应用于基因的快速定位外，还可以用于构建高分辨率的遗传图谱，QTL作图(Wang 和 Paterson 1994)以及长期选择后代的分群分析 (Soghroue 和Rocheford 1994) 。

在 n 个F_2个体所组成的混合物中至少发生一次双交换的概率如图5.14所示。对于5 cM的区间，当混合物所含个体数为40或更少时，此概率小于0.10 (10%)。 当目标区间增大到10cM时，混合个体数不到10即可保持相同的双交换概率。但是随着样本数的减少，两类混合物间在除目标区间以外的区域保持同质的机会就会大大增加，从而导致PCR检测时假阳性的增加。因此，Giovannoni等(1991)建议混合体所含个体数应大于5。每一类混合体含5个个体，其结果是97%的非目标基因组区域将来自于双亲。他们还认为目标区间的长度应小于15cM。

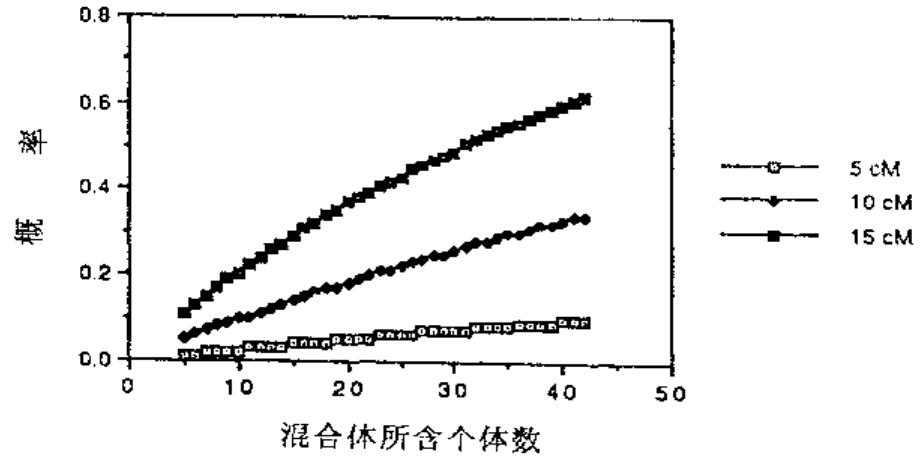


图5.14 在n个F_2个体所组成的混合体中至少在目标区间内发生一次双交换的概率目标区间的长度分别为5cM，10cM和15cM，选择目标区域相邻的标记纯合的F_2组成混合体

第 6 章　QTL作图：均值-方差分析法

以RFLP 等分子标记为基础的技术可以帮助育种学家更好地了解数量性状的遗传基础，并为定位和操纵与数量性状表达有关的单个基因座位提供有效的途径，特定 QTL 的鉴别将为上位性、多效性及杂种优势遗传基础的研究提供更有力的工具，已定位遗传标记的应用将大大加快植物数量性状改良和选择的进展。通过QTL 分子作图、物理图谱制作等，还可实现对 QTL 的分子克隆和遗传操纵。本书第6～9章将介绍各种 QTL 作图方法的原理及其应用。

完整的分子标记图谱的建立，使得有可能对品系间数量遗传差异起作用的多个孟德尔因子进行分解。在原理上，对多基因性状的剖分包括：(1)由目标性状有差异的品系产生分离群体；(2)对群体中各个体或品系进行该性状及在基因组上分布的共显性标记进行考查记载；(3)寻找性状与一个或多个标记遗传模式之间的相关。所采用的方法主要是以标记为基础的 MB 法和以性状为基础为TB法(徐云碧 1992)。一旦发现了显著的相关，即认为存在一个 QTL，由此将多基因分解为单个的QTL。

MB 法可以分成两类，一类一次只考虑一个标记座位，另一类一次考虑所有标记座位并在对其它座位进行校正的同时估计每一座位的效应。对于从两个纯合品系所产生的后代，在回交、RIL 和 DH群体中，只有两种标记基因型可用于比较，因此可以对标记基因型均值进行t测验或方差分析的F测验。对于 F_2或 F_3家系，有三种可能的标记基因型，既有纯合体又有杂合体，也可只利用标记纯合体均值进行比较。 以性状为基础的方法是上述一次一个标记方法的变形(Lebowitz等1987)。在以性状为基础的分析中，从供试群体分布的尾部抽取个体或品系来计算标记频率。两尾间频率有差异的标记被认为与影响该性状的QTL有关。

自从Sax(1923)报道质量性状与数量性状之间存在相互关联以来，Harrison和Mather (1950)、Thoday(1961) 以及 Wehrhahn和Allard(1965)对 QTL 的检测和作图进行了探讨。根据与经典的形态学和细胞学标记之间的连锁关系，Thoday (1961)提出了定位与数量性状有关的染色体区段或QTL 的方法。此后，许多学者讨论了借助于当时可供利用的遗传标记进行QTL定位的理论方面 (Elston和Steward 1973; Hill 1975; Jayakar1970; Mather和Jinks 1971; McMillan和 Robertson 1974; Soller 等1976)。Thoday 等人的理论是建立在分离群体中标记基因型之间数量性状平均值差异的基础上的。 这种以标记为基础(Marker-Based) 的平均值差异法(简称MB 法) 是80年代以前及80年代初期利用遗传标记进行 QTL 分析的主要方法。建立在标记与数量性状之间相互关联基础上的另一类方法是 Stuber等(1980)提出的以性状为基础 (Trait-Based method) 的方法 (简称TB法)。许多学者根据不同群体的遗传特性，分别提出了相应的标记座位与QTL相互关联的检测方法，所涉及的群体包括：近交系间分离群体(Soller和Beckmann 1983; Edwards等1987; Asins和 Carbonell 1988; Hu和 Stuber1992)，异交系间分离群体 (Beckmann和Soller 1988)，

加倍单倍体 (DH) 群体(Snape 1988)，两个近交系间的重组近交系 (RIL) (Simpson 1989) 以及一粒传群体(SSD) (Snape 1988)。

在理论上，检测关联或连锁时所采用的统计方法都涉及到标记基因型均值的应用。但本章所介绍的方法是直接利用这类均值和方差所进行的分析，不仿称为均值-方差分析法。

6.1 基本假定

假定某一特定的分子标记座位M具有两个等位基因 M_1 和 M_2 (互为共显性)，特定的数量性状基因座位 Q具有两个等位基因 Q_1 和 Q_2，两近交亲本 P_1 和 P_2 的基因型分别为 $M_1M_1\ Q_1Q_1$ 和 $M_2M_2Q_2Q_2$，标记座位M与QTL之间的重组率为 r。则 F_2 群体中三类标记基因型 M_1M_1、M_1M_2、M_2M_2，DH 群体中两类标记系 M_1M_1、M_2M_2，以及 F_1 分别与 P_1 和 P_2 回交的BC群体中两类标记基因型 M_1M_2、M_2M_2 和 M_1M_1、M_1M_2 的平均值和方差，F_2 群体中 QTL 基因型 Q_1Q_1、Q_1Q_2、Q_2Q_2，DH 群体中 Q_1Q_1、Q_2Q_2，以及BC群体中 Q_1Q_2、Q_2Q_2 和 Q_1Q_1、Q_1Q_2 的平均值和方差，可假定如表6.1、表6.2和表6.3所示。

以下分别用下标 i，j 和 k 表示第 i 类标记基因型，第 j 类QTL基因型，第 i 类标记基因型中的第 k 个体或品系，而第 i 类标记基因型所含个体(品系)数为 n_i。因此，对所有 i 和 j，F_2：$i,j=1,\ 2,\ 3$；DH (BC)：$i,j=1,\ 2$；对于所有的 k，$k=1,\ 2,\ \ldots,\ n_i$。

为后面讨论的方便，我们将 F_2 和DH (BC) 群体中各类基因型的频率以矩阵的形式分别表示为

表6.1 F_2 群体中标记座位M和QTL各基因型的频率及平均值和方差

基因型	Q_1Q_1	Q_1Q_2	Q_2Q_2	样本均值	样本方差	样本容量
M_1M_1	$(1-r)^2$	$2r(1-r)$	r^2	m_1	s_1^2	n_1
M_1M_2	$r(1-r)$	$(1-2r+2r^2)$	$r(1-r)$	m_2	s_2^2	n_2
M_2M_2	r^2	$2r(1-r)$	$(1-r)^2$	m_3	s_3^2	n_3
平均值	μ_1	μ_2	μ_3			
方　差	σ_1^2	σ_2^2	σ_3^2			

表6.2 DH 群体中标记座位M和QTL各基因型的频率及平均值和方差

基因型	Q_1Q_1	Q_2Q_2	样本均值	样本方差	样本容量
M_1M_1	$(1-r)$	r	m_1	s_1^2	n_1
M_2M_2	r	$1-r$	m_2	s_2^2	n_2
平均值	μ_1	μ_2			
方　差	σ_1^2	σ_2^2			

表6.3A 与P_1回交的BC群体(B_1)中标记座位M和QTL各基因型的频率及平均值和方差

基因型	Q_1Q_1	Q_1Q_2	样本均值	样本方差	样本容量
M_1M_1	$(1-r)$	r	m_1	s_1^2	n_1
M_1M_2	r	$1-r$	m_2	s_2^2	n_2
平均值	μ_1	μ_2			
方　差	σ_1^2	σ_2^2			

表6.3B 与P_2回交的BC群体(B_2)中标记座位M和QTL各基因型的频率及平均值和方差

基因型	Q_1Q_2	Q_2Q_2	样本均值	样本方差	样本容量
M_1M_2	$(1-r)$	r	m_1	s_1^2	n_1
M_2M_2	r	$1-r$	m_2	s_2^2	n_2
平均值	μ_1	μ_2			
方　差	σ_1^2	σ_2^2			

$$p_{ij}=\begin{bmatrix}(1-r)^2 & 2r(1-r) & r^2\\ r(1-r) & 1-2r+2r^2 & r(1-r)\\ r^2 & 2r(1-r) & (1-r)^2\end{bmatrix}\quad i,\ j=1,\ 2,\ 3$$

和

$$p_{ij}=\begin{bmatrix}1-r & r\\ r & 1-r\end{bmatrix}\quad i,\ j=1,\ 2$$

6.2 标记-性状关联的检测

6.2.1 回归分析法

如果将作图群体每一个体的标记基因型用数据形式表示，例如用1和0来区别来自两个近交系的回交群体中的两种标记基因型(纯合体和杂合体)，用2、1和0来区别两个近交系之间的 F_2群体中的三种标记基因型(纯合体、杂合体和另一纯合体)，就可以据此判断标记基因是否是数量性状存在关联。

将标记与数量性状变异联系起来的最为简单的方法就是测验特定标记不同标记类型的个体间性状值的差异 (Soller等1976)。以回交群体为例，如果我们假定 m_1和 m_2是回交群体中标记基因型为 M_1M_1 和M_1M_2的两类个体的数量性状均值，我们就可以利用通常的

t测验来检验均值 m_1 和 m_2 之间的差异显著性。检验的假设可以是：

H_0: $m_1 = m_2$

和 H_1: $m_1 \neq m_2$

在统计上，这等价于具有下列模型的简单回归分析

$$y_j = b_0 + b_1 x_j + \varepsilon_j \qquad j=1,\ 2,\ \ldots,\ N$$

其中，y_j 为群体中第 j 个体的性状值，b_0 为模型的均值(回归截距)，x_j 为第 j 个体的标记类型，当标记基因型为 M_1M_1 时取值为1，为 M_1M_2 时取值为0。b_1 为简单回归系数，ε_j 为第 j 个体的随机剩余变量，假设具有正态分布，

$$\varepsilon_j \sim N(0,\ \sigma^2)。$$

对于回交群体，$N = n_1 + n_2$。在该模型中可对回归系数进行检验。

假定存在 q 个 QTL对两个近交品系间回交群体的遗传变异起作用。从遗传上讲，若忽略上位性，则 m_1 和 m_2 之间的期望差异为

$$E(m_1 - m_2) = \sum_{i=1}^{q} (1-2r_i)(d_i + h_i)$$

其中 E 表示期望，d_i 和 h_i 分别为轮回亲本的纯合体和杂合体之间差异的第 i 个 QTL的加性效应和显性效应，r_i 为第 i 个QTL与标记的重组率。这意味着我们所检验的是由许多基因的基因效应和重组率所组成的一个复合参数。当然，许多 QTL可能不与该标记相连锁，并且因而重组率为0.5。因此上述假设检验等价于

H_0: 所有 $r_i = 0.5$ 和 H_1: 至少有一个 $r_i < 0.5$

因为我们假定 $d_i + h_i$ 是非零的（即通过试验设计我们知道在群体中有一些基因处于分离状态)。如果发现 m_1 和 m_2 之间存在显著的差异，则表明该标记与一个或可能多个QTL相连锁。

如果假定在某特定染色体上最多只存在一个与标记有关的QTL，则上述假设检验等价于

H_0: $r = 0.5$ 和 H_1: 至少有一个 $r < 0.5$

在不作特别说明的情况下，我们均假定在特定染色体上只存在一个可能与标记相连锁的QTL。

回归方程的解为

$$\hat{b}_1 = \frac{\sum (x_j y_j) - (\sum x_j \sum y_j)/N}{\sum x_j^2 - (\sum x_j)^2 / N}$$

$$= \frac{\sum (x_j - \bar{x})(y_j - \bar{y})}{\sum (x_j - \bar{x})^2}$$

$$= SCP / SS_x$$

其中 SCP 为依变量 Y 和自变量 X 的乘积和，SS_x 是自变量 X 的平方和。

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x}$$

用回归系数的估计值 $\hat{b}_0$ 和 $\hat{b}_1$ 以及自变量 X 可以估计依变量 Y：

$$\hat{y}_j = \hat{b}_0 + \hat{b}_1 x_j$$

$\hat{y}_j$ 与 $\hat{y}$ 之差称为剩余值，剩余值平方的总和称为剩余平方和，记为 SSE，

$$SSE = \Sigma\ (y_j - \hat{y}_j)^2$$

剩余均方 MSE 为

$$MSE = SSE\ /\ (N-2)$$

对于简单回归分析，我们可以利用方差分析的方法将依变量 Y 的总变异分解为对自变量 X 的线性回归所引起的变异和随机误差引起的剩余变异，即总平方和 SST 可以分解为回归平方和 SSR 和剩余平方和 SSE：

$$SST = SSR + SSE$$

其中 $SSR = \Sigma(\hat{y}_j - \bar{y})^2 = (\hat{b}_1)^2 SS_x$

而回归均方为 $MSR = SSR\ /1 = SSR$

当无效假设 H_0: $b_1=0$ 成立时，$F = MSR/MSE$ 具有自由度为(1, $N-2$)的 F 分布。因此可用 F 测验对回归系数进行显著性测验(表6.4)。

表6.4 简单线性回归的方差分析表

变异来源	自由度	平方和	均方	期望均方	F值
线性回归	1	SSR	MSR	$\sigma^2 + b_1^2$	$MSR\ /\ MSE$
离回归	$N-2$	SSE	MSE	σ^2	
总和	$N-1$	SST			

回归平方和 SSR 与总平方和 SST 之比定义为决定系数(Coefficient of determination)，一般用 r^2 表示，即

$$r^2 = SSR\ /\ SST$$

它度量了依变量 Y 对自变量 X 的线性回归所能解释的变异占总变异的比例，在标记-性状关联分析中，就是与标记相关联QTL所解释的变异占总变异的比例，它反映了这种关联或连锁的 QTL效应的贡献大小。

回归系数 $\hat{b}_1$ 的显著性也可以采用t测验进行检验。

b_1 的估计值 $\hat{b}_1$ 是平均数为 b_1、方差为 σ^2/SS_x 的正态变量，因而

$$(\hat{b}_1 - b_1)\ /\ \sigma(\hat{b}_1)$$

具有标准正态分布。在实际应用时只能获得 $\hat{b}_1$ 的样本方差 $\hat{\sigma}^2(\hat{b}_1) = MSE\ /\ SS_x$。$(\hat{b}_1 - b_1)\ /\ \hat{\sigma}(\hat{b}_1)$ 具有自由度为 $N-2$ 的 t 分布。因此可用 $t_{(N-2)}$ 分布对回归系数进行显著性检验。

6.2.2 单向分组方差分析法

如果标记座位与 QTL无连锁，则标记与数量性状在遗传上不存在相关，亦即 F_2 群体

各类标记基因型之间的数量性状平均值彼此相等， $m_1=m_2=m_3$。将作图群体中不同标记基因型的个体类型看作独立样本，则可根据每类样本容量不相等的单向分组方差分析来判断各类标记基因型之间数量性状的差异。该方差分析的线性模型为：

$$y_{ij}=\mu+\tau_i+\varepsilon_{ij}$$

其中：y_{ij}=第 i 类基因型中第 j 个体的数量性状表型值

μ =作图群体数量性状总均值

$\tau_i=m_i-\mu$

$\varepsilon_{ij}=y_{ij}-m_i$

因而具有表 6.5 所示的方差分析表。其中：

$$df_T=\sum n_i-1$$

$$df_t=k-1$$

$$df_e=df_T-df_t$$

$$SS_T=\sum\sum y_{ij}^2-(\sum\sum y_{ij})^2/\sum n_i$$

$$SS_t=\sum\frac{(\sum y_{ij})^2}{n_i}-\frac{(\sum\sum y_{ij})^2}{\sum n_i}$$

$$SS_e=SS_T-SS_t$$

$$n_0=\frac{(\sum n_i)^2-\sum n_i^2}{(\sum n_i)\cdot(k-1)}$$

表 6.5　不同标记基因型之间数量性状的单向分组方差分析

变异来源	df	SS	MS	EMS
基因型之间	df_t	SS_t	MS_t	$\delta_e^2+n_0\delta_\tau^2$
误　差	df_e	SS_e	MS_e	δ_e^2
总变异	df_T	SS_T		

6.2.3　均值和方差比较法

当考虑标记座位 M和 QTL时，双亲间 F_2和 DH或 BC群体中分别有 9种和4种基因型，它们的相对频率为重组值 r 的函数(假定 Q_1Q_1、Q_1Q_2和 Q_2Q_2的基因型值分别为 d、h 和 $-d$，其中 d 为基因加性效应，h 为基因显性效应)。将每一标记基因座位基因型的频率乘以Q座位的基因型值并求和，得到不同标记基因型的期望均值，并根据其频率求出其相应的期望方差如表6.6、6.7和表6.8，其中 μ 作图群体数量性状的总均值。

根据不同标记基因型之间数量性状的期望平均数和方差，可知在 F_2中，当 $r=0$（即标记座位本身控制着数量性状而表现为多效性)时，标记基因型 M_1M_1，M_1M_2和M_2M_2的均值分别简化为Q座位 Q_1Q_1，Q_1Q_2 和 Q_2Q_2 的基因型值 $\mu+d$，$\mu+h$ 和 $\mu-d$，它们是不等的，而其方差则是相等的。当 $r=0.5$时，3种标记基因型的均值是相等的，均为 $\mu+h/2$，方差则是不等的。而在$0<r<0.5$的范围内，各标记基因型的均值是不相等的，

表 6.6　F_2群体标记基因型内数量性状的期望均值和方差

基因型	期望均值	期望方差
M_1M_1	$\mu+d(1-2r)+2hr(1-r)$	$2d^2r(1-r)+2h^2rs-4dhr(1-3r+2r^2)$
M_1M_2	$\mu+h(1-2r+2r^2)$	$2d^2r(1-r)+2h^2rs$
M_2M_2	$\mu-d(1-2r)+2hr(1-r)$	$2d^2r(1-r)+2h^2rs+4dhr(1-3r+2r^2)$

$s=1-3r+4r^2-2r^3$

表 6.7　DH群体标记基因型内数量性状的期望均值和方差

基因型	期望均值	期望方差
M_1M_1	$\mu+d(1-2r)$	$4d^2r(1-r)$
M_2M_2	$\mu-d(1-2r)$	$4d^2r(1-r)$

表 6.8　与P_1和P_2回交的BC群体中标记基因型内数量性状的均值和方差

回交方式	标记基因型	期望均值	期望方差
与P_1回交	M_1M_1	$\mu+d(1-r)+hr$	$(d-h)^2r(1-r)$
	M_1M_2	$\mu+dr+h(1-r)$	$(d-h)^2r(1-r)$
与P_2回交	M_1M_2	$\mu-dr+h(1-r)$	$(d+h)^2r(1-r)$
	M_2M_2	$\mu-d(1-r)+hr$	$(d+h)^2r(1-r)$

方差也是不相等的，且标记基因型为 M_1M_1的个体群具有最小的方差，M_2M_2个体群有最大的方差。

在DH 群体中，当 $r=0$时，M_1M_1 和 M_2M_2 之间的期望均值是不等的，而方差是相等的，且有关的遗传方差等于0。当 $r=0.5$时，均值是相等的，方差也是相等的。当$0<r<0.5$时，均值是不等的，而方差是相等的。

在BC群体中，当 $r=0$时，M_1M_1和 M_1M_2之间、M_2M_2和 M_1M_2之间的期望均值是不等的，而方差是相等的，且与标记有关的遗传方差等于0。当 $r=0.5$时，均值是相等的，方差也是相等的。当$0<r<0.5$时，均值是不等的，而方差是相等的。

因此，可以根据各标记基因型之间的观测平均数和方差的大小及期望的遗传方差与环境方差的异质性检验，判断标记座位与QTL的连锁关系。其中环境方差可由双亲的平均方差估计。

根据以上原理，利用上述群体中不同标记基因型的均值和方差可以构建检验标记-性状关联的所需的有关统计量。Zhang等(1992)讨论了与之有关的一些问题，我们在引述时作了一些补充。

1. 利用F_2群体标记基因型内数量性状均值的差异

当 $r=0.5$ 时

$$m_1-m_3=2d(1-2r)=0$$

$$m_1-m_2=d(1-2r)-h(2r-1)=0$$

$$m_2-m_3=-d(1-2r)-h(2r-1)=0$$

上述差数的标准误为

$$\sqrt{s_i^2/n_i+s_j^2/n_j} \qquad (i,j=1,2,3;\quad i\neq j)$$

因此可用 t 测验来检验标记与假定存在的 QTL之间是否独立遗传，其自由度为

$$df=\frac{(s_i^2/n_i+s_j^2/n_j)^2}{\dfrac{(s_i^2/n_i)^2}{n_i-1}+\dfrac{(s_j^2/n_j)^2}{n_j-1}} \qquad (i,j=1,2,3;\quad i\neq j)$$

2. 利用F_2群体标记基因型内数量性状的方差

由标记基因型内个体间的方差可以构成如下方差比

$$\frac{4s^2}{s_1^2+2s_2^2+s_3^2}=\frac{d^2/2+h^2/4+V_e}{[1-(1-2r)^2]d^2/2+[1-(1-2r)^4]h^2/4+V_e}$$

$$=\frac{1}{1-[(1-2r)^2+(1-2r)^4D^2/2]h_N^2}$$

其中 s^2 为 F_2群体数量性状的表型方差，$D=h/d$ 为显性度，h_N^2为狭义遗传率。

当$0\leqslant r\leqslant 0.5$时，上述方差比将$\geqslant 1$；当 $r=0.5$时，方差比为1。如果三种标记基因型各自组成的亚群体是相互独立的，则方差比近似地服从自由度为 df_1和df_2的 F 分布，其中 $df_1=df_{F2}$，

$$df_2=\frac{(s_1^2+2s_2^2+s_3^2)^2}{\dfrac{(s_1^2)^2}{n_1-1}+\dfrac{4(s_2^2)^2}{n_2-1}+\dfrac{(s_3^2)^2}{n_3-1}}$$

3. 利用相关系数

分别给 F_2群体中三种标记基因型 M_1M_1、M_1M_2、M_2M_2 以特定的编码值(如1、2、3)，则群体中各个体的标记基因型编码值与相应的数量性状值之间的相关系数为

$$R=(1-2r)\sqrt{d^2/(2s^2)}$$

R 的标准误为

$$S_R=\sqrt{(1-R)/(N-2)}$$

$df=N-2$，当 R 的估值与0有显著差异时即表明标记与 QTL连锁，反之亦然。在统计学上，这与回归分析的原理相类似。

6.3 重组率及QTL遗传参数的估计

标记基因型的均值和方差可用于获得两个独立的重组率估计值，同时估计出QTL的有关遗传参数(加性效应和显性效应)。

6.3.1 利用均值估计

根据数量遗传理论，存在

$$P_1 - P_2 = 2d \tag{1}$$

$$m_1 - m_3 = 2(1-2r)d \tag{2}$$

并且定义

$$L_1 = \left| \frac{m_1 - m_3}{P_1 - P_2} \right| = (1-2r) \tag{3}$$

如果三种标记基因型的亚群体相互独立，则 L_1 的方差为

$$V_{L_1} = \frac{1}{(P_1-P_2)^2}\left[\frac{s_1^2}{n_1-1} + \frac{s_3^2}{n_3-1}\right] + \frac{(m_1-m_3)^2}{(P_1-P_2)^4}\left[\frac{s_{P1}^2}{df_{P1}} + \frac{s_{P2}^2}{df_{P2}}\right]$$

由(3)式得重组率的估计值为

$$\hat{r} = (1-L_1)/2$$

相应的方差为

$$V\hat{r} = V_{L_1}/4$$

将 r 的估计值代入(2)即得基因加性效应的估计值

$$\hat{d} = \frac{m_1 - m_3}{2(1-2\hat{r})} = \frac{m_1 - m_3}{2L_1}$$

其方差为

$$V_{\hat{d}} = \frac{1}{4L_1^2}\left[\frac{m_1}{n_1-1} + \frac{m_3}{n_3-1}\right]$$

显性效应的估计式为

$$\hat{h} = \frac{m_2 - (m_1 + m_3)/2}{(1-2\hat{r})^2}$$

或

$$\hat{h} = \frac{m_2 - (m_1 + m_3)/2}{L_1^2}$$

相应的方差为

$$V_{\hat{h}} = \frac{1}{(1-2\hat{r})^4}\left\{\frac{s_2^2}{n_2-1} + \frac{1}{4}\left[\frac{s_1^2}{n_1-1} + \frac{s_3^2}{n_3-1}\right]\right\}$$

或

$$V_{\hat{h}} = \frac{1}{L_1^4}\left\{\frac{s_2^2}{n_2-1} + \frac{1}{4}\left[\frac{s_1^2}{n_1-1} + \frac{s_3^2}{n_3-1}\right]\right\}$$

由于可能存在的QTL同效等位基因在双亲间分散分布的影响，两个亲本的表型差值并不能真实地反映作图群体中存在的真实遗传差异。此时可考虑 F_2 群体中表型极端的两类个体。当群体足够大时，F_2 群体将分离出不同 QTL 上同效等位基因集中分布的个体，增效和减效等位基因的纯合体就是群体中的极端表型个体。因此可以从群体中选择这两类个

体来代替亲本 P_1和 P_2进行上述分析，以估计重组率、基因加性效应和显性效应。有关公式中的亲本均值和方差分别用两类极端类个体的均值和方差代替。根据基因分离理论，如果假定群体中只存在一个 QTL 的分离，则在 F_2群体中增效纯合体和减效纯合体所占比例均为25%。由于数量性状的表现型易受环境的干扰，因此在实际选择时，只需从中选择10株左右的极端个体(假定 F_2群体大小在60以上)。

6.3.2 利用方差估计

利用回交群体标记基因型的方差可得：

$$[\ s^2_{B1(M1M1)} + s^2_{B1(M1M2)}\ ] - [\ s^2_{B2(M1M2)} + s^2_{B2(M2M2)}\ ]$$
$$= -8r(1-r)dh \quad \text{记为} \ s^2_{BC}$$

其中B_1和B_2表示分别为与P_1和P_2回交。利用F_2群体标记基因型的方差可得

$$s^2_1 - s^2_3 = -8r(1-r)(1-2r)dh \quad \text{记为} \ s^2_F$$

s^2_F和s^2_{BC}之比记为L_2，则

$$L_2 = |\ s^2_F \ / \ s^2_{BC}\ | = \frac{8r(1-r)(1-2r)\quad dh}{8r(1-r)\,dh} = (1-2r)$$

其方差为

$$V_{L2} = \frac{2}{(s^2_{BC})^2}\left[\frac{(s^2_1)^2}{n_1-1} + \frac{(s^2_3)^2}{n_3-1}\right]$$
$$+ \frac{2(s^2_F)^2}{(s^2_B)^4}\left[\frac{[s^2_{B1(M1M1)}]^2}{df_{B1(M1M1)}} + \frac{[s^2_{B1(M1M2)}]^2}{df_{B1(M1M2)}}\right.$$
$$\left. + \frac{[s^2_{B2(M1M2)}]^2}{df_{B2(M1M2)}} + \frac{[s^2_{B2(M2M2)}]^2}{df_{B2(M2M2)}}\right]$$

因此，重组率的估计值为

$$\hat{r} = (1-L_2)/2$$

相应的方差为

$$V\hat{r} = V_{L2}/4$$

可以采用 t 测验来检验重组率的显著性。类似于利用均值估计重组率时的方法也可获的QTL加性效应和显性效应的估计值。

6.3.3 连锁检测所需的最小植株数

如果 F_2群体中两种标记等位基因纯合体的数目完全符正常的分离比例，则 $n_1 = n_3$ 设该值为 N'。为了在5%的显著水平、II型错误的概率为 0.10($z_{\alpha/2}$=1.96, z_β=1.28)的条件下检测到连锁的存在，则下述条件一定得到满足：

$$3.24 \leqslant (m_1 - m_3) \ / \ \sqrt{(s^2_1 + s^2_3)/N'}$$

其中

$$m_1 - m_3 = 2d(1-2r)$$

若不考虑环境方差，则

$$s^2_1 + s^2_3 = 4d^2r(1-r) + 4h^2r(1-3r+4r^2-2r^3)$$

设数量性状的显性度为

$$D = h / d$$

则上述条件等式可以重写为

$$N' \geqslant 10.5[r(1-r)+D^2 r(1-3r+4r^2-2r^3)]/(1-2r)^2$$

由此可以得到满足上述条件的最小 N' 值。图6.1给出了当显性度 D=0, 1/4, 1/2, 3/4, 1和1.5时不同重组率下的最小 N' 值。类似地，也可以将标记基因杂合体用于这种分析。

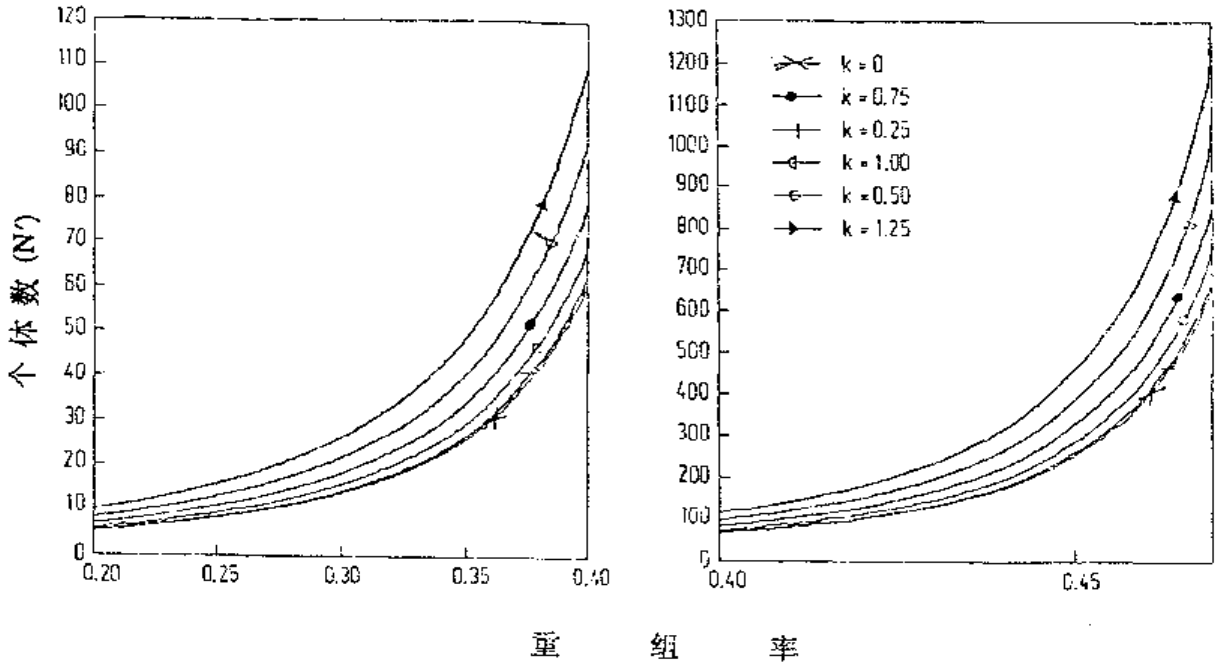


图 6.1　在5%的显著水平、0.1的II型错误概率下检测连锁所需的最小样本数 (N')

在不存在基因型×环境互作的假定下，可以将环境方差引入上述分析中。根据广义遗传率的定义可得环境方差为

$$V_e = (d^2/2 + h^2/4)(1/h^2-1)$$

假定在不同标记基因型内，环境方差保持不变，则

$$s_1^2+s_3^2 = 4d^2 r(1-r)+4h^2 r(1-3r+4r^2-2r^3)+d^2(1+D^2/2)(1/h^2-1)$$

进而得到

$$N \geqslant 10.5[r(1-r)+D^2 r(1-3r+4r^2-2r^3)+0.25(1+k^2/2)(1/h^2-1)]/(1-2r)^2$$

故此时每种标记纯合体样本容量的增加量为

$$2.625(1+k^2/2)(1/h^2-1)/(1-2r)^2$$

因此

当 $0 \leqslant D \leqslant 1.25$ 时，$1 \leqslant (1+k^2/2) \leqslant 1.8$；

当 $0 \leqslant r \leqslant 1/3$ 时，$1 \leqslant 1/(1-2r)^2 \leqslant 9$；

当$0.5 \geqslant h^2 \geqslant 0.1$时，$1 \leqslant (1/h^2-1) \leqslant 9$。

在 D、r 和 h^2 三个参数的最差组合下，每种纯合体将需要额外增加383个。

6.4 利用标记基因频率检验连锁

6.4.1 基本原理

以性状为基础的分析方法的基本原理是，由于选择的结果改变了QTL中处于分离状态的增效或减效等位基因的频率，使数量性状的高表型个体中QTL增效等位基因以及低表型个体中减效等位基因的频率增加。QTL 等位基因与其相邻的标记等位基因之间的“免费搭车旅行”效应将使标记等位基因的频率发生相应的变化。当QTL 等位基因与某一标记基因连锁时，由于相互关联的结果，必然导致高低表型个体间标记基因频率的差异。因此，在高低两类选系中基因频率具有显著差异的标记座位被认为与QTL连锁在一起(Lebowitz等(1987)。根据Falconer (1981)，当显性效应 $h=0$ 时，在F_2和BC群体中进行一轮双向选择的结果，QTL 基因频率的变化分别为：

$$S_Q = k f / 4 \qquad (F_2)$$

$$S_Q = k f / 8 \qquad (BC)$$

其中 $f = 2d / \sigma_g$ 是QTL 的比例效应，k 为选择强度，σ_g 为群体基因型标准差。经过 t 轮选择，QTL基因频率的变化近似于 tS_Q。QTL基因频率的变化使与之连锁的标记基因频率发生相应的变化，在1轮选择和 t 轮选择时分别为

$$S_{m1} = (1-2r) S_Q$$

和

$$S_{mt} = \sum_{j=1}^{t} (1-2r)(1-r)^{j-1} S_Q$$

6.4.2 试验设计

Lebowitz等(1987)提出了三种试验设计以用于这类以性状为基础的QTL分析。

设计 I：用于分析在胁迫环境下 F_2或BC群体中存活的个体。测定存活个体中的标记基因的频率，并与期望的孟德尔频率 (F_2: 0.5; BC: 0.75 和0.25)进行比较。同时在未遭受胁迫选择的群体中对基因频率与期望值有显著差异的标记基因进行分析，以排除遗传标记异常分离的可能性。这类设计特别适合于耐盐性、耐旱性、耐寒性等耐逆性与标记基因的关联分析。

设计 II：用于分析从两个近交系间杂种所选出的品系。对于单向选择所得品系，比较标记基因频率与 F_2的期望频率。对于双向选择所得的高表型和低表型两类品系，比较相互间标记基因频率的差异。这类设计便于与育种目标相结合。

设计 III：从 F_2 或BC群体中挑选两类极端表型的个体进行分析。这类设计在对标记考查所需成本明显高于子代培育及对数量性状考查所需成本时采用，因而有利于减少工作量，降低分析成本。

TB法的优点在于它特别适合于与育种和选择试验相结合。尽管实际应用极少报道，但作为这种方法的延伸，在利用分子标记定位质量基因方面，先根据质量性状的表型分群，

寻找与之有关的分子标记，然后利用分离群体加以验证，可使分析工作量大大减小。

6.5 应用与实例

6.5.1 应用概况

Sax (1923)第一个报道了形态标记性状与数量性状的相互关联。他发现菜豆种子的大小(数量性状)与种皮色素（离散的单基因性状）之间存在着有色与大粒、白色与小粒的相互关联。他把这种相关归结于控制种子大小的遗传因子与种皮色素基因的连锁。不过这种关联现象也有可能是种皮色素基因的多效性所致。Sax 的试验没有排除多效性的可能性。Rasmusson (1935) 在研究豌豆开花期的变异时发现了开花期与色素的连锁。以其中一个组合为例，有色花品种Gj的平均开花期(与以若干品种平均开花期的离差表示)为8.5天，白色花品种Bism为 -9.3天，有色花对白色花决定于一对主基因 *A-a*，这两个品种的 F_2群体中出现了类似于Sax (1923) 所观察到的现象，即有色花植株的平均开花期迟于白色花植株，且差异达极显著水平(表6.9)。但在此之前，曾在该组合中选出了一个早开花的白色花品系 HRT-II (平均开花期为-6.1天)。在 HRT-II×Bism的 F_2群体中，开花期的差异与颜色有关的现象消失了。而在 HRT-II 与迟开花的白色花品种 St的 F_2代，开花期与花色的关系发生了逆转，有色花植株的开花期显著早于白色花植株。这种关系的逆转正是交换打破连锁的必然结果。因此，原组合 Gj×Bism 中开花期与色素基因的关联是控制开花期的多基因与控制色素的主基因相互连锁的结果，而不是由于色素基因的多效作用。这是采用形态标记检测与之关联的QTL 的经典例子。

表 6.9 豌豆开花期与花色的遗传

组 合	F_2植株的平均开花期		开花期的差异
	有色(*A* –)	无色 (*aa*)	(*A* – – *aa*)
Gj × Bism	5.37 ± 0.31	2.11 ± 0.76	3.26
HRT-II×Bism	-7.97 ± 0.36	-8.30 ± 0.81	0.33
HRT-II×St	-1.24 ± 0.20	1.63 ± 0.23	-2.87

Everson和Schaller(1955) 利用大麦品种 Atlas 与以 Atlas为轮回亲本育成的高产品系Calif 1306杂交，研究了与刺芒有关的产量差异的遗传。刺芒主基因 (*R-r*)和穗轴毛长度主基因 (*S-s*)位于第五连锁群。Atlas表现刺芒、短毛穗轴，基因型为 *RRss*，产量较低。Calif1306 为半光芒、长毛穗轴，基因型为 *rrSS*，产量较高。在这两个品系的F_2中，分离出四种纯合基因型 *RRSS*、*RRss*、*rrSS*、*rrss*。从基因型为 *RrSs* 的F_2植株中分离出大量的 *RRSS* 和 *rrss* 型的F_3株系。由于高产与刺芒基因(*r*)有关，大多数 *rrss* 品系的产量将高于 *RRSS* 品系，如果真实的情形与此相反，则表明刺芒基因与产量

基因间发生了交换，且增加产量的基因为位于 *Rs*/ *rS* 的片段之内。 反之，如果 *rr*/ *ss* 品系间或 *RR* /*ss* 品系间不存在产量差异，则表明 *r* 基因存在多效性或与产量基因紧密连锁或产量基因位于 *rs* 片段之外。8个F_2基因型为 *rs*/ *rs* 的 F_8品系，株系间产量的均方显著，表明这些品系间存在真实的产量差异。其中两个品系的产量极显著低于群体平均值，三个品系极显著地高于群体平均值，说明这两类品系间具有不同的产量基因。两个低产品系的产量与亲本 Atlas相当，它们可能缺乏高产的等位基因，这些品系来源于 *R-r* 座位与产量基因座位间的交换，因而带有于 Atlas 相同的降低产量的等位基因。 这些结果提供了高产与刺芒的联系被打破的证据，表明这种联系是由于基因连锁而不是基因多效性所致。

在动物方面，Thoday及其同事以果蝇为材料，借助于带有遗传标记的测验品系，在X染色体及第2、3染色体上分别鉴别出影响腹部刚毛数的数量遗传因子 (Spikett和 Thoday 1966;Thoday1961)。其中5个基因控制着高刚毛品系和野生品系间刚毛数目差异的87.5%。

就细胞遗传学标记而言，在小麦上曾成功地利用非整倍体及其衍生系进行QTL的定位。由于单体和端着丝点染色体出现在小麦的每一对染色体中，采用供体亲本的单个染色体去代换受体亲本的对应染色体，产生了完整的21 个品种间代换系，使每一个品种的每一对染色体分别被一共同亲本的对应染色体所取代。 这样两个品种间的遗传差异就可以一个一个染色体的分割开来，而且每个代换系就象一个品种一样，成为一个稳定的品系，因而可以重复种植每一品系，更精确地研究具有数量特征的性状。 Sears 及其同事分别以小麦品种中国春 (CS) 和希望 (H) 为受体和供体，培育出含有希望品系染色体的整套中国春代换系 CS(H)。虽然这两个品种都是春小麦，但希望的抽穗期早于中国春，品种间 F_2抽穗期一般表现为连续变异，且显著超过迟抽穗的亲本。 将代换系 CS(H) 与中国春、希望种植于15°C恒温连续光照的控制条件下，发现其中10个系抽穗期显著早于或迟于中国春；其中 CS(H5A)、CS(H5D) 和 CS(H7B) 三个代换系与 CS 的抽随期相差最大。其中希望的5A和7B带有早熟的等位基因，而5D 则带有迟熟的等位基因 (Law 和Gale 1979)。

上述品种间染色体代换系所反映的可能是代换染色体与受体的对应染色体之间大量基因差异的总效应。为了确定这些差异的特性，一种方法是将代换系与受体亲本杂交，产生纯合背景下单个染色体的杂合体 (Law 1966)。该杂种衍生物之间差异就反映了两个对应染色体有差异的基因的分离。 通过对这些杂种衍生物的研究，可以估计所定位的基因数目及其效应。如果将单染色体杂种与所研究的染色体为单体的受体品种回交，则它们的衍生物为该染色体的半合重组体或非重组体，将这些衍生物自交，可选择到稳定遗传的单个染色体的杂合体，该杂合体可用于检测品系（而不是株系）间的差异。将这种方法进一步扩展，可用来产生一个代换染色体与受体品种中同源的端着丝点染色体相结合的杂种，使代换染色体只有一个臂是杂合的，而另一个臂是半合的，因此交换只会发生在一个臂上。对该杂种衍生物的研究可以发现基因所在的染色体臂及其与着丝点的相对距离。Law 等将 CS 双着丝点 $7B^L$与CS(H7B) 杂交形成一个臂 ($7B^L$) 杂合、一个臂 ($7B^S$) 半合的杂合体 (交换只会发生在$7B^L$上)。将该杂合体再与CS单体(7B) 杂交，根据测交后代性状是否发生分离来判断 QTL所在的染色体臂位 (图6.1)。结果表明，Law (1966) 所证实的控制生育期的遗传因子位于第 7B染色体的短臂上，因为杂种后代生育期的双峰分布只出现在 $7B^S$ 的

衍生物中，该因子距离着丝点25±6cM。对CS(H5A)和CS(H5D)所进行的类似研究表明，在相应染色体的长臂上存在着控制生育期的另外两个基因（Law 1966）。

以标记基因型之间的表型均值为基础，采用同工酶作为遗传标记检测标记-数量性状关联的研究主要是有关玉米和番茄的。在玉米中已证实了同工酶标记与产量基因的连锁（Stuber 等1980）。在番茄中，Tanksley等采用以标记为基础的平均值差异法，利用回交群体，确定了与柱头外露性、叶片长宽比、果重和种子重四个数量性状有关的QTL及其所在的染色体。Weller 等(1988)研究了番茄中6个形态标记和4个同工酶标记与18个数量性状的相互关联。在大豆中，Graef等(1989)研究了两个 *G. Soja* × *G. max* 回交群体中同工酶标记与数量性状之间的关联。

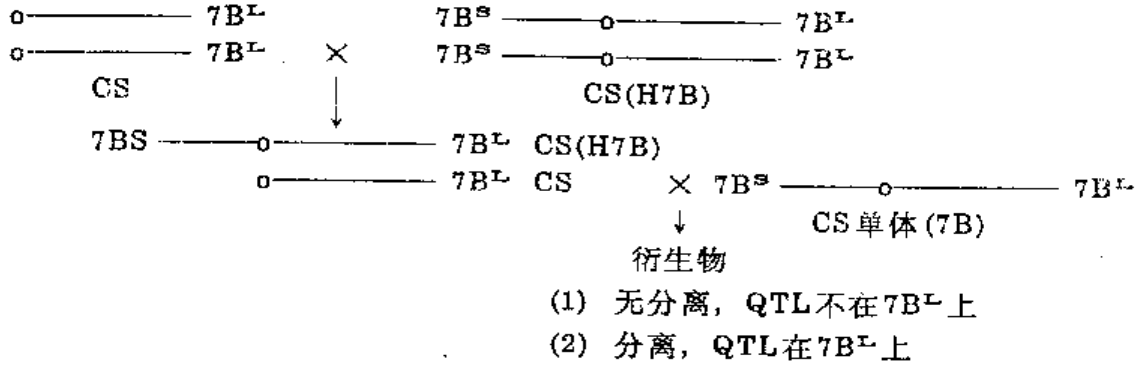


图 6.2 利用端着丝点染色体确定QTL所在臂位

利用 RFLP作为遗传标记，在番茄中现已报道了 RFLP 与抗虫性 (Nienhuis等1987)、可溶性固形物含量 (Osborn等1987)以及水分利用率 (Martin等1989)等数量性状基因之间的关联。在大豆中，分子标记已用于鉴别 *Glycine max* × *G. Soja* 衍生群体的QTL。Keim等(1990a)利用70个 RFLP 标记研究了 *G. max* × *G. soja* 单交群体中的种子硬实性，所发现的 QTL解释了群体硬实性总变异的71%。Keim等(1990b) 以同一群体继续作图，利用150个 RFLP 标记发现了与成熟期和形态性状有关的 QTL，且可以解释20%以上的性状总变异。Diers等(1992) 采用同一群体，利用243个 RFLP、5个同工酶、1个贮藏蛋白和三个形态学的标记，进行了种子蛋白质和含油量的基因型分析，发现对于特定性状，分离标记可以解释的总变异有的高达43%。Fatokun 等(1992)采用RFLP 图谱定位了影响豇豆和绿豆种子重量的QTL，其中两个非连锁基因组区域所发现的QTL解释了豇豆种子重量变异的52.7%。在玉米中，Beavis等(1991)利用4个 F_2 群体研究了影响株高性状的QTL，所鉴定的QTL大多位于已定位的质量遗传基因座位附近。Edwards等(1992)利用98个 RFLP 探针和16个同工酶标记考查了过去利用同工酶作为遗传标记研究过的玉米 F_2 群体中标记基因与数量性状之间的关联，发现与QTL 关联的 RFLP 标记座位与过去的有关结果十分一致。Reiter等(1991)在控制试验下评价了玉米自交系间90个 F_3 家系在低磷胁迫下的营养生长，发现77个RFLP标记中有6个与低磷胁迫下的表现存在显著关联，其中一个标记基因座位可以解释25%的总表型变异，所有鉴别的 QTL均以加性基因作用为主，其中两个标记与加性

×加性的上位性互作相关联，包括3个标记基因座位和显著上位性互作的多元回归模型解释了总表型变异的46%。在大麦中，Powell等(1992)以rDNA的基因间隔顺序为多态性指标，以大麦DH群体为材料，发现了rDNA等位基因与大麦产量、千粒重、水分敏感性(反映种子休眠性)以及籽粒加工能量等数量性状的显著关联。

近年来，有关分子标记与与数量性状相互关联的报道很多，在此不一一列举。

6.5.2. 数据实例

作者以籼型水稻窄叶青8 (ZYQ)和粳型水稻京系17 (JX) 为亲本配制杂交组合，F_1自交产生 F_2群体，利用分子标记构建了连锁遗传图谱(见图4.3)。该作图群体由76个植株组成，共调查了10个农艺性状。两个亲本各10个植株的数量性状列于表6.10。F_2群体5个数量性状和第5染色体上的7个RFLP标记座位的基因型数据分别列于表6.11和表6.12。我们以此将作为后续有关章节中QTL作图的数据实例。同时我们也欢迎有兴趣的同行将其应用于各自的研究中。

表6.10　产生作图群体的亲本ZYQ和JX的性状表现

株	窄叶青 8 (ZYQ)					京系 17 (JX)				
号	生育期	株高	穗长	颖花数	结实率	生育期	株高	穗长	颖花数	结实率
1	106	78	18.1	132	83.33	101	109	20.7	181	92.82
2	106	91	19.9	171	84.21	104	112	23.4	173	88.44
3	108	91	23.5	212	79.72	104	108	20.6	161	83.23
4	105	90	21.4	217	81.11	104	105	21.7	163	79.14
5	106	96	20.0	233	84.12	101	103	21.0	127	81.89
6	107	91	20.7	215	80.47	102	109	22.1	172	80.23
7	105	93	19.5	170	81.18	102	107	20.5	120	90.83
8	106	91	20.2	231	79.65	106	105	19.5	141	90.07
9	108	89	19.7	186	77.42	104	102	20.0	112	89.29
10	108	92	21.5	234	77.35	104	108	20.6	145	88.28
均值	106.5	90.2	20.5	200.1	80.86	103.2	106.8	21.0	149.5	86.42

1. 标记-性状关联的检测

利用标记基因型之间数量性状的平均数比较以及单向分组方差分析方法，发现了分布于第5染色体上的 7 个 RFLP 标记与5个数量性状之间的关联。其中均值比较发现了16个显著或极显著的比较/性状组合，涉及到所有的标记和性状(表6.13)。单向分组方差分析发现了8个显著的标记/性状组合，涉及到除生育期以外的4个性状和除RG474 和RG435 以外的5个标记，其中RZ70/穗长、RG70/颖花数、RG697 /结实率三个组合的关联达极显著水平(表6.14)。RG573 也同时与两个相关性状(穗长和颖花数)存在显著关联，这表明同一标记同时与多个相关性状存在着显著的相互关联。

表 6.11　水稻窄叶青8和京系17亚种间F_2群体的5个数量性状表型值

株号	性状					株号	性状				
	生育期	株高	穗长	颖花数	结实率		生育期	株高	穗长	颖花数	结实率
1	126	115	20.5	129	42.64	39	124	107	21.7	224	69.64
2	107	127	27.5	356	44.94	40	109	106	25.4	254	60.63
3	112	123	29.0	341	25.81	41	106	123	25.2	186	60.22
4	95	117	20.5	202	60.89	42	106	108	20.4	115	78.26
5	115	121	24.0	285	26.67	43	92	100	24.2	260	82.69
6	110	127	24.3	311	34.73	44	115	112	20.5	188	52.66
7	101	103	20.7	126	42.06	45	113	113	24.0	270	61.11
8	119	116	25.2	237	46.84	46	108	113	25.5	242	15.70
9	118	112	22.0	217	56.68	47	125	106	21.9	158	44.94
10	115	111	24.7	186	55.91	48	120	105	19.5	126	39.68
11	104	115	27.5	330	28.79	49	101	110	22.5	216	34.72
12	109	106	22.5	162	59.88	50	121	104	23.5	161	79.50
13	107	115	26.0	270	40.00	51	124	91	21.5	189	16.93
14	114	123	23.7	314	44.27	52	122	110	21.5	182	43.96
15	123	97	20.7	146	49.32	53	112	111	21.5	186	77.42
16	111	111	24.9	254	35.04	54	114	127	23.1	180	84.44
17	119	110	22.0	322	72.05	55	106	97	18.1	162	27.16
18	113	126	28.8	357	28.57	56	124	122	24.2	214	72.90
19	124	102	25.7	223	18.39	57	89	109	23.0	172	47.67
20	102	110	23.0	231	45.02	58	112	110	23.2	198	56.57
21	116	126	24.8	310	21.94	59	114	140	27.0	202	73.27
22	113	111	19.5	201	47.26	60	118	94	20.0	187	86.10
23	115	132	28.3	411	27.49	61	106	110	22.7	176	81.82
24	101	102	22.8	171	32.16	62	117	106	25.0	227	36.56
25	109	120	23.4	233	36.05	63	114	96	22.2	154	13.64
26	111	138	26.5	294	61.90	64	111	107	25.5	220	76.36
27	110	122	29.9	224	60.27	65	119	116	24.5	189	43.92
28	115	123	27.5	233	47.21	66	109	90	21.5	145	23.45
29	114	111	24.0	197	28.93	67	108	95	22.1	149	64.43
30	109	119	26.5	237	39.66	68	104	120	22.5	167	52.10
31	114	119	24.5	221	49.77	69	103	101	20.5	134	76.12
32	119	109	20.0	273	37.00	70	99	104	23.5	225	80.00
33	105	126	30.6	245	28.98	71	109	124	25.5	210	48.10
34	113	134	27.0	327	55.05	72	115	118	28.7	237	59.49
35	119	110	25.7	360	80.56	73	112	118	22.3	198	30.30
36	116	123	24.0	217	71.43	74	111	135	28.9	213	49.77
37	107	118	23.4	183	64.48	75	114	100	19.8	155	6.45
38	114	123	25.5	207	51.69	76	117	131	26.5	327	44.34

表6.12　水稻窄叶青8和京系17 亚种间F_2群体的RFLP标记基因型

株号	RFLP标记							株号	RFLP标记						
	RG573	RG13	RZ70	RG697	RG474	RG435	RZ225		RG573	RG13	RZ70	RG697	RG474	RG435	RZ225
1	1	1	3	1	1	1	1	39	3	3	3	2	2	1	3
2	2	2	2	3	3	3	3	40	2	2	2	2	2	2	2
3	2	1	1	2	1	1	2	41	2	3	2	2	2	2	2
4	1	1	3	1	2	1	1	42	1	3	3	2	3	3	1
5	3	3	1	3	2	3	3	43	2	2	3	2	2	1	1
6	2	2	1	2	1	2	2	44	1	3	3	2	3	3	1
7	3	3	2	2	1	2	2	45	2	2	2	2	2	2	2
8	2	2	2	0	3	3	3	46	2	2	2	1	2	2	1
9	1	2	3	0	2	2	3	47	2	1	2	2	2	2	2
10	3	3	3	0	3	3	3	48	3	3	2	1	1	1	1
11	2	3	3	0	2	2	2	49	2	2	2	2	2	2	2
12	2	2	3	2	2	2	2	50	2	2	3	2	1	1	2
13	2	2	2	0	2	2	2	51	2	2	2	2	2	2	2
14	3	3	2	3	3	3	3	52	3	3	3	2	3	3	3
15	1	1	3	0	1	1	2	53	3	3	3	2	2	2	2
16	3	3	2	3	2	3	3	54	3	3	3	2	1	1	2
17	2	3	2	2	2	2	2	55	3	3	3	3	3	3	3
18	1	0	2	2	2	0	1	56	1	1	2	2	2	1	2
19	2	1	1	3	2	3	3	57	3	3	3	3	3	3	3
20	2	2	1	0	2	2	2	58	2	2	2	2	2	2	1
21	2	2	2	1	2	2	1	59	1	1	2	2	2	2	2
22	2	2	2	2	3	3	3	60	2	2	2	3	3	3	3
23	2	3	2	2	2	3	0	61	1	1	1	2	2	2	2
24	3	2	2	1	3	3	3	62	3	2	3	3	3	3	3
25	1	1	2	1	0	1	1	63	3	2	3	3	2	2	3
26	2	1	2	0	2	1	2	64	3	3	3	3	3	3	3
27	1	1	2	1	1	2	1	65	1	1	1	1	1	1	1
28	2	2	2	2	2	2	2	66	3	3	3	3	3	3	3
29	2	3	2	2	2	2	2	67	1	1	2	1	1	1	1
30	2	2	3	3	3	3	3	68	2	2	2	2	2	2	2
31	2	2	2	2	1	2	2	69	3	3	3	2	2	2	2
32	2	3	2	2	2	2	2	70	2	2	2	3	3	3	3
33	2	2	2	3	3	3	2	71	2	2	3	3	3	3	3
34	2	2	2	2	2	2	2	72	1	2	1	0	1	1	1
35	3	3	2	2	2	2	2	73	3	3	3	3	2	2	2
36	3	2	2	2	2	2	2	74	2	2	2	1	2	2	2
37	3	2	3	2	2	1	2	75	1	1	2	3	3	3	3
38	2	2	2	2	2	2	2	76	1	2	1	1	1	1	1

1: ZYQ等位基因，2:杂合体，3：JX等位基因

表6.13　ZYQ/JX F_2群体标记基因型之间数量性状均值之差的 t 检验

标记	均值之差比较	数量性状				
		生育期	株高	穗长	颖花数	结实率
RG573	m_1-m_2	2.21	-0.19	-1.06	-41.74*	6.55
	m_1-m_3	2.19	5.99	0.96	3.26	3.69
	m_2-m_3	-0.02	6.18*	2.03**	45.00**	-2.86
RG13	m_1-m_2	3.84	0.24	-0.66	-28.32	0.51
	m_1-m_3	2.87	3.63	1.10	-16.71	-1.11
	m_2-m_3	-0.97	3.39	1.76**	11.62	-1.63
RZ70	m_1-m_2	1.02	1.78	1.00	20.92	-5.73
	m_1-m_3	2.49	8.44*	2.75**	68.93**	-11.84
	m_2-m_3	1.47	6.67*	1.75**	48.01**	-6.11
RG697	m_1-m_2	-1.04	1.65	0.46	-14.98	-13.67*
	m_1-m_3	1.50	6.75	0.23	-10.92	3.00
	m_2-m_3	2.54	4.90	-0.24	4.07	16.67*
RG474	m_1-m_2	3.05	-0.79	-0.05	-31.95	1.45
	m_1-m_3	5.10*	4.63	0.87	-1.96	4.12
	m_2-m_3	2.05	5.42	0.92	29.98	2.67
RG435	m_1-m_2	2.35	0.15	-0.40	-14.44	8.12
	m_1-m_3	2.98	3.97	-0.01	-9.70	13.84*
	m_2-m_3	0.63	3.82	0.40	4.75	5.73
RZ225	m_1-m_2	-0.72	-1.40	-0.26	-7.13	-4.38
	m_1-m_3	-0.77	5.97	0.73	2.92	5.48
	m_2-m_3	-0.06	7.38*	0.99	10.04	9.86

*和**表示分别达5%和1%显著水平

表6.14　ZYQ/JX F_2群体标记基因型之间数量性状的单向分组方差分析

标记	F测验自由度		数量性状				
	df_1	df_2	生育期	株高	穗长	颖花数	结实率
RG573	2	73	0.56	2.48	4.15*	4.83*	0.66
RG13	2	72	1.37	0.81	3.26*	1.09	0.05
RZ70	2	73	0.48	3.75*	5.24**	6.88**	1.48
RG697	2	64	0.67	1.66	0.14	0.23	5.45**
RG474	2	72	1.93	1.67	0.78	2.18	0.20
RG435	2	72	0.83	1.00	0.20	0.29	2.61
RZ225	2	72	0.06	3.52*	0.94	0.20	1.84

*和**表示分别达5%和1%显著水平

方差比分析表明，凡平均数比较和单向分组方差分析显著的标记-性状组合，其方差比值均大于1(表6.15)，但均未达显著水平。方差比值与单向分组方差分析中的 F 值以及均值比较中的 t 值大小成正比，这表明利用F_2群体总方差和各标记基因型方差所构造的方差比的确可以用于揭示标记-性状间可能存在的相互关联。

标记基因型的编码值与相应的数量性状之间的相关分析表明，有四个标记-性状的关联达到显著或极显著水平(表6.16)，这些组合是：RZ70-株高(穗长、颖花数)，RG435-结实率。其中RZ70与穗长和颖花数这两个相关性状的关联极为显著，这与方差分析和均值比较的结果完全一致。

表6.15　ZYQ/JX F_2群体标记基因型之间数量性状方差比分析

标记	F测验自由度		数量性状				
	df_1	df_2	生育期	株高	穗长	颖花数	结实率
RG573	75	72	0.993	1.025	1.047	1.098	1.001
RG13	74	64	0.999	0.944	1.033	1.078	0.976
RZ70	75	63	1.055	1.096	1.131	1.125	1.019
RG697	66	57	0.966	1.007	0.937	0.987	1.174
RG474	74	72	1.043	1.009	0.951	0.999	1.005
RG435	74	59	0.994	0.996	1.004	0.994	1.058
RZ225	74	67	0.974	1.066	0.970	0.949	1.035

方差比为 $$F = \frac{4s^2}{s_1^2+2s_2^2+s_3^2}$$

表 6.16　ZYQ/JX F_2群体标记基因型与数量性状表型之间的相关系数

标记	df	生育期	株高	穗长	颖花数	结实率
RG573	74	-0.098	-0.207	-0.151	-0.049	-0.057
RG13	73	-0.109	-0.131	-0.186	0.068	0.025
RZ70	74	-0.113	-0.287*	-0.349**	-0.386**	0.197
RG697	65	-0.084	-0.211	-0.018	0.045	-0.104
RG474	73	-0.223	-0.158	-0.117	-0.021	-0.073
RG435	73	-0.139	-0.142	-0.008	0.048	-0.258*
RZ225	73	0.034	-0.222	-0.111	-0.026	-0.124

*和**表示分别达5%和1%显著水平

根据数量性状的表现将F_2群体分成高于F_2均值和低于F_2均值的两个亚群，分析了亚群间各标记等位基因频率的差异(表6.17)。结果表明，对于大部分标记，其ZYQ等位基因

表6.17　ZYQ/JX F_2群体数量性状高低表现型(高于F_2均值即$>M$和低于F_2均值即$<M$)个体间ZYQ等位基因的频率

标记		生育期	株高	穗长	颖花数	结实率
RG573	$>M$	0.477	0.557	0.500	0.485	0.514
	$<M$	0.455	0.390	0.434	0.453	0.427
RG13	$>M$	0.429	0.529	0.473	0.438	0.443
	$<M$	0.439	0.354	0.395	0.430	0.425
RZ70	$>M$	0.419	0.471	0.500	0.530	0.329
	$<M$	0.348	0.317	0.276	0.279	0.439
RG697	$>M$	0.461	0.517	0.453	0.426	0.516
	$<M$	0.448	0.405	0.457	0.475	0.403
RG474	$>M$	0.512	0.544	0.487	0.469	0.500
	$<M$	0.391	0.390	0.432	0.453	0.425
RG435	$>M$	0.488	0.544	0.432	0.438	0.557
	$<M$	0.409	0.378	0.474	0.465	0.363
RZ225	$>M$	0.452	0.544	0.473	0.469	0.529
	$<M$	0.455	0.378	0.434	0.442	0.387
F_2性状均值M		111.9	113.7	23.9	222.9	49.4

的频率在0.40～0.55，群体中籼型等位基因的频率偏低（低于理论值0.5)。但对于有些标记-性状组合，在两个亚群体间，ZYQ 等位基因的频率差异较大。其中频率之差大于 0.15 的有8个标记-性状组合，而大于0.20的只有 RZ70-穗长和 RZ70-颖花数两个，这与上述其它分析所得结果基本类似。

标记与数量性状的相互关联是标记基因座位与控制该性状的QTL相互连锁所造成或是由于标记基因的一因多效。而同一标记与多个相关性状的关联则可用于解释数量性状之间可能存在的遗传相关。一般认为造成遗传相关的重要原因是基因连锁或一因多效，即控制不同数量性状的 QTL相互连锁或同一QTL 影响两个或两个以上数量性状的遗传。要区分这两种可能的原因，还有待于 QTL的精确作图。

在理论上，当标记基因型之间数量性状的平均值不相等时，标记-性状之间的关联可能是由于标记基因的一因多效（$r=0$）而不是连锁（$0<r<0.5$)。而当方差不等时，标记-性状之间可能存在连锁（$0<r<0.5$）也可能是相互独立的（$r=0.5$)。据此可以判断标记和数量性状间相互关联的遗传原因。在本例中，均值比较中关联极显著的标记/性状组合均得到方差分析结果的证实，表明这些极显著的关联是由于标记-QTL连锁的结果。

根据标记基因型与数量性状表现型之间的相互关联，只能检测到正好处于标记座位的 QTL，其有效性在很大程度上取决于遗传图谱的饱和程度及同一分离群体所容纳的遗传标记的多少。过去由于形态标记在数量上的限制以及部分标记表型上的有害性，通过标记与

数量性状之间的关联所能发现的 QTL 是十分有限的。RFLP 等分子标记由于其数量多且表型上呈中性，可以在一个作图群体中定位大量的标记，并使之覆盖全部基因组，因而在理论上有可能检测到与某一数量性状有关的所有 QTL。徐云碧等(1994)采用位于水稻12条染色体上的52 个 RFLP 标记(覆盖了水稻的大部分基因组)，通过标记基因型之间数量性状的平均值和方差比较及单向分组方差分析，共揭示出显著关联的标记-性状组合共9个，即 RG101-株高、RG544-株高和颖花密度、RG449-分蘖角、RG573-穗长、RG172-株高和颖花密度、 RG1-分蘖角、RG869-颖花密度，并认为这些标记-性状关联可能是由于标记基因与控制相应数量性状的 QTL 之间连锁的结果。随着遗传图谱中已定位的分子标记的逐渐增加，这类方法的效率也将不断提高，并使之成为可以普遍接受而又简单易行的方法。

2. 标记-QTL重组率及QTL基因效应的估计

根据 F_2群体标记基因型的均值及亲本均值所估计的标记-QTL重组率及相应的基因加性效应和显性效应列于表6.18。重组率的估计值分布于 -1.966～0.494之间。在正的重组率估计值中，经 t 测验，均与独立时的期望重组率(0.50)无显著差异。 由于在估计重组率时采用了两亲本的表型均值之差，因此负的重组率估计值可能与双亲间数量性状的表型差异较小有关，而决定亲本表型差异大小的因素除了受遗传差异的大小影响之外，还与它们之间 QTL等位基因的分布有关，一对存在遗传差异的亲本可能由于同效等位基因的分散分布而具有相同的表型(详见第11章)。其结果使亲本间的表型差异小于 F_2群体中两类标记纯合体的差异，而导致负的重组率估计值，并同时影响基因显性效应和加性效应的估计。此外，在理论上，重组率估计值愈接近0.50，其相应的加性效应和显性效应的估计值就愈高。因此，只有当重组率的估计值与0.50存在显著差异时，基因效应的估值才是有意义的，表6.18所列数据供与下述分析比较时参考。

为克服 QTL同效等位基因在亲本间的分散分布对重组率及相应的基因效应估计的影响，我们采用 F_2群体的极端表现型个体来估计原始亲本中不能表现为表型的真实差异。结果表明，所有标记-性状组合的重组率估计值均处于0.00～0.50的理论值范围内 (表6.19)。经 t 测验，有五个标记-QTL 的连锁达显著水平，即重组率估计值与0.50存在显著差异，其中RZ70与穗长和颖花数的 QTL连锁达极显著水平。不显著的标记-性状组合，其重组率估值均在0.40以上并接近0.50。

对于显著的标记-QTL 连锁，估计了有关的基因效应(表6.20)。t 测验表明，加性效应的显著性与其重组率的显著性相平行，其中影响穗长和颖花数的 QTL的加性效应达极显著水平，可分别使穗长增加4.6cm，每穗颖花数增加108个。基因显性效应估计值均未达显著水平。

3. 与标记连锁的QTL 所能解释的变异占总变异的比例

在回归分析中，不仅可以根据 F 测验和 t 测验的结果判断标记与性状的关联是否达到显著水平，而且可以将数量性状的总变异分解为线性回归所引起的变异和随机误差所引起的剩余变异，决定系数 (r^2)所反映的就是与标记相关联的QTL所解释的变异占总变异的比例。对ZYQ/JX F_2群体的分析表明，F 测验或 t 测验显著的标记-性状组合共有8个，其中4个极显著。对于显著的标记-性状组合，标记所能解释的变异占性状总变异的比

例为 1.23%～14.91%。 其中效应最大的 RZ70分别能解释穗长和颖花数变异的12.15%和 14.91%。

表 6.18 根据 ZYQ/JX F_2群体标记基因型的均值和方差估计的标记-QTL重组率(r)、基因加性效应(d)和显性效应(h)

标记	参数	生育期	株高	穗长	颖花数	结实率
RG573	r	0.168± 0.399	0.320± 0.110	-0.360± 1.263	0.468± 0.204	0.166± 0.614
	d	1.650± 1.952	8.300± 4.956	0.280± 0.275	25.300±160.051	2.783± 5.071
	h	-2.534± 4.005	24.445±19.914	0.522± 0.217	--	-10.705±10.392
RG13	r	0.066± 0.437	0.391± 0.124	-0.483± 1.393	0.335± 0.225	0.400± 0.640
	d	1.650± 1.624	8.300± 9.339	0.280± 0.253	-25.300± 33.800	-2.783±17.830
	h	-3.187± 2.425	33.002±68.678	0.312± 0.168	183.194±129.081	-26.850±119.20
RZ70	r	0.123± 0.464	0.246± 0.115	-1.956± 2.806	-0.181± 0.296	-0.564± 0.796
	d	1.650± 2.005	8.300± 3.634	0.280± 0.094	25.300± 8.510	-2.783± 1.875
	h	0.400± 3.194	9.458±10.217	0.016± 0.027	7.297± 8.457	-0.042± 1.100
RG697	r	0.273± 0.484	0.297± 0.139	0.299± 1.118	0.392± 0.236	0.231± 0.638
	d	1.650± 3.499	8.300± 5.610	0.280± 1.528	-25.300± 54.917	2.783± 6.536
	h	8.644± 9.693	9.229±17.383	-2.162± 4.604	204.657±359.243	52.334±16.282
RG474	r	-0.273± 0.418	0.361± 0.122	-0.274± 1.351	0.481± 0.242	0.130± 0.600
	d	1.650± 0.828	8.300± 7.226	0.280± 0.381	-25.300±315.915	2.783± 4.420
	h	-0.209± 0.733	39.941±34.353	0.202± 0.290	--	1.107± 8.411
RG435	r	0.048± 0.451	0.381± 0.115	0.494± 0.846	0.404± 0.214	-0.743± 0.677
	d	1.650± 1.612	8.300± 7.901	-0.280±37.314	-25.300± 56.118	2.783± 1.215
	h	-1.054± 2.212	32.166±46.407	--	261.257±406.242	-0.193± 0.727
RZ225	r	0.383± 0.431	0.320± 0.101	-0.151± 1.141	0.471± 0.213	0.008± 0.610
	d	-1.650± 6.054	8.300± 4.546	0.280± 0.384	25.300±186.661	2.783± 3.328
	h	5.956±33.101	33.903±19.482	0.368± 0.393	--	7.350± 4.683

-- ：略去了由于 r 估计值接近0.5而获得的偏大的 h 估计值。
负的 r 估计值可能由于两亲本间数量性状的差异较小所致(与同效等位基因的分布状态有关)

表 6.19 根据ZYQ/JX F_2群体标记基因型的均值和方差估计的标记-QTL重组率(r)（估计式中的亲本均值和方差由F_2群体中的两类极端表型个体的均值和方差的代替，极端类型个体数=8)

标记	生育期	株高	穗长	颖花数	结实率
RG573	0.459	0.421	0.448	0.493	0.471
RG13	0.446	0.452	0.440	0.462	0.491
RZ70	0.453	0.389*	0.351**	0.341**	0.407
RG697	0.472	0.411	0.488	0.475	0.477
RG474	0.404*	0.439	0.453	0.495	0.468
RG435	0.444	0.448	0.500	0.478	0.392*
RZ225	0.485	0.421	0.461	0.493	0.457

表6.20 根据ZYQ/JX F_2群体标记基因型的均值和方差估计的标记-QTL重组率(r)、基因加性效应(d)和显性效应(h)及其标准误
(只列出重组率显著的标记-性状组合，估计式中的亲本均值和方差由F_2群体中的两类极端表型个体的均值和方差的代替，极端类型个体数=8)

标记-性状组合	参数估值及其标准误		
标记 - 性状	$r \pm s_r$	$d \pm s_d$	$h \pm s_h$
RZ70 株高	0.389±0.049	19.000± 8.318	49.562± 53.541
RZ70 穗长	0.351±0.050	4.619± 1.552	4.237± 7.314
RZ70 颖花数	0.341±0.054	108.687±36.560	134.665±156.071
RG474 生育期	0.404±0.049	13.250± 6.648	-13.467± 47.263
RG435 结实率	0.392±0.047	31.941±13.943	-25.453± 95.716

表6.21 ZYQ/JX F_2群体数量性状表现型与标记基因型的回归分析
(F为方差分析的F值，t为回归系数显著性检验t值，r^2为决定系数)

标记	参数	生育期	株高	穗长	颖花数	结实率
RG573	F	0.7205	3.3108	1.7319	0.1761	0.2409
	t	0.7023	1.0363	2.9917**	0.0404	0.1555
	r^2	0.0096	0.0428	0.0229	0.0024	0.0032
RG13	F	0.8847	1.2717	2.6130	0.3441	0.0470
	t	0.7740	0.6347	3.7532**	0.0579	0.0687
	r^2	0.0120	0.0171	0.0346	0.0047	0.0006
RZ70	F	0.9514	6.6590*	10.2385**	12.9637**	2.9993
	t	0.7260	1.3483	6.8906**	0.3372	0.5019
	r^2	0.0127	0.0826	0.1215	0.1491	0.0390
RG697	F	0.4582	3.0393	0.0205	0.1307	0.7176
	t	0.4819	0.8570	0.2799	0.0298	0.2239
	r^2	0.0070	0.0447	0.0003	0.0020	0.0109
RG474	F	3.8248	1.8745	1.0134	0.0332	0.3887
	t	1.5292	0.7160	2.1068*	0.0162	0.1834
	r^2	0.0498	0.0250	0.0137	0.0005	0.0053
RG435	F	1.4427	1.5005	0.0042	0.1699	5.2228*
	t	1.0134	0.7054	0.1513	0.0415	0.7651
	r^2	0.0194	0.0201	0.0001	0.0023	0.0668
RZ225	F	0.0855	3.7677	0.9099	0.0501	1.1434
	t	0.2386	1.1177	2.1712*	0.0226	0.3400
	r^2	0.0012	0.0491	0.0123	0.0007	0.0154

第 7 章　QTL作图：矩法和最大似然法

标记基因和QTL之间重组率及其有关遗传参数的准确估计对于以标记为基础的选择是极为重要的。这是因为标记选择的效率一方面取决于与QTL 相连锁的标记所能解释的加性方差的比例，另一方面也取决于标记-QTL 连锁估计的准确性。因此，在将分子标记应用于数量性状的选择之前，必须首先确定与数量性状有关的遗传标记及其与 QTL的重组率。

采用适当的统计方法，利用F_2或回交群体等的资料可以准确定位QTL与标记座位之间的重组率，并确定QTL 不同等位基因的遗传参数。Zhuchenko等 (1979)试图利用矩量法解决回交群体中的这类问题，但未能恰当地测验其方法的有效性。Jayakar (1970)曾为两类不同试验设计提出了标记-QTL 之间重组率的估计式，但未能给出环境方差的适当估计方法。Weller (1986)在数量性状呈正态分布的假定下，将矩量法与最大似然法相结合，利用两个近交系间 F_2群体及各类标记基因数量性状统计量来估计遗传标记与 QTL 之间的重组率以及各类 QTL 基因型的平均值和方差，从而在理论上开创了利用最大似然法确定QTL与标记座位之间重组率的先例。对于这一方法，共显性基因比显性基因更为有效，但它不能保证其解就是真正的最大似然估计值。而且该方法需要花费大量的计算时间，需要借助于计算机工作站才能完成其分析。Luo和Kearsey(1989，1991)采用基本上与Weller(1986)类似的方法通过数学上的处理使最大似然估计只涉及重组率一个参数，发展了分别适合于近交F_2、回交(BC) 群体和加倍单倍体 (DH) 群体的统计方法。并认为其方法适合于遗传率不小于0.1的单个 QTL 控制的数量性状，但对于同样大小的遗传率而被多个QTL 等分时，该方法是无效的。Darvasi和Weller (1992)认为该方法只是一种近似最大似然估计。Luo和Woolliams (1993) 将其改称为矩估计法，并发展了相应的适合于近交F_2群体的最大似然估计法。同时还考虑了QTL基因型间数量性状方差的同质性(方差同质模型) 和异质性(方差异质模型)。

本章介绍在上述基础上发展完善的以单一标记为基础、估计标记-QTL 重组率的矩估计法和最大似然法。

7.1　原理与统计方法

对于作图群体中分子标记座位 M 和特定的数量性状基因座位 Q，采用与第6章完全相同的假定。两近交亲本 P_1和 P_2之间的 F_2 群体、BC以及DH 群体中标记基因型的样本均值 m_i 和方差 s_i^2，以及QTL基因型的平均值 μ_j和方差 σ_j^2 见表6.1和6.2，且有用矩阵形式表示的各类基因型的频率 p_{ij}。

如果标记与 QTL之间的连锁是不完全的，则各标记基因型内的个体应包括有各种QTL基因型。这些数量性状基因型具有平均值 μ_j。在数量性状基因型之间具有不同方差（方差异质模型）的假定下，其方差为 σ_j^2。其中 σ_j^2 包含环境的变异以及影响该数

量性状的其它基因座位所引起的遗传变异。在数性状基因型之间具有相同方差（方差同质模型）的假定下，$\sigma_j^2 = \sigma^2$。各类标记基因型的样本均值 m_i 和样本方差 s_i^2 可以分解为

$$m_1 = p_{12}\mu_1 + p_{12}\mu_2 + p_{13}\mu_3 \tag{1}$$

$$s_i^2 = \sum_{j=1} p_{ij}[\sigma_j^2 + (\mu_j - \overline{m_i})^2] \tag{2}$$

如果进一步假定各QTL基因型中的个体，其观测值服从平均数为 μ_j、方差为 σ_j^2 的正态分布，则对于标记基因型 M_i (F_2: $i=1,2,3$; DH (BC): $i=1,2$) 内部的分布就是几种数量性状基因型的正态分布的混合体，其密度函数可以表示为各正态分布按其比例加权的线性组合 (Day 1969; Hasselbald 1966):

$$f_{Mi}(x) = \sum_{j=1} p_{ij} f_j(x)$$

其中:

$$f_j(x) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left[-\frac{(x-\mu_j)^2}{2\sigma_j^2}\right]$$

因此似然函数可写作:

$$L = L(\mu_j, \sigma_j^2) = \prod_{i=1}^{} \prod_{k=1}^{n_i} f_{Mi}(x_{ik})$$

两边取对数得:

$$\ln L = \sum_{i=1} \sum_{k=1}^{n_i} \ln f_{Mi}(x_{ik})$$

$$= \ln \frac{1}{\sqrt{2\pi}} \sum_{i=1} n_i + \sum_{i=1} \sum_{k=1}^{n_i} \ln\left\{\sum_{j=1} \frac{p_{ij}}{\sigma_j} \exp\left[-\frac{(x_{ik}-\mu_j)^2}{2\sigma_j^2}\right]\right\} \tag{3}$$

方程(3)中除了 r 外，还含有未知参数 μ_j 和 σ_j^2。

7.1.1 矩估计法

Luo和Woolliams(1993)给出了F_2群体中未知参数向量 $\Phi=(r, \mu_j, \sigma_j^2)$ 中 μ_j 和 σ_j^2 的矩估计式。徐云碧(1994a)将其扩展到DH和BC群体。

如果定义 $d_{ij}=(\mu_j - m_i)^2$

$$e_i = s_i^2 - \sum_{j=1} p_{ij} d_{ij}$$

则对于F_2群体，方程组(1)和(2)的解为:

$$\hat{\mu}_1 = \frac{1}{(1-2r)^2}\{(1-r)^2 m_1 - 2r(1-r)m_2 + r^2 m_3\}$$

$$\hat{\mu}_2 = \frac{1}{(1-2r)^2}\{r(r-1)m_1 - [(1-r)^2 + r^2]m_2 + r(r-1)m_3\}$$

$$\hat{\mu}_3 = \frac{1}{(1-2r)^2}\{r^2 m_1 - 2r(1-r)m_2 + (1-r)^2 m_3\}$$

$$\hat{\sigma}_1^2 = \frac{1}{(1-2r)^2}\{(1-r)^2 e_1 - 2r(1-r)e_2 + r^2 e_3\}$$

$$\hat{\sigma}_2^2 = \frac{1}{(1-2r)^2}\{r(r-1)e_1 - [(1-r)^2 + r^2]e_2 + r(r-1)e_3\}$$

$$\hat{\sigma}_3^2 = \frac{1}{(1-2r)^2}\{r^2 e_1 - 2r(1-r)e_2 + (1-r)^2 e_3\}$$

则对于DH (BC) 群体，方程组(1)和(2)的解为:

$$\hat{\mu}_1 = \frac{1}{(1-2r)}[(1-r)m_1 - r\,m_2]$$

$$\hat{\mu}_2 = \frac{1}{(1-2r)}[-r\,m_1 + (r-1)m_2]$$

$$\hat{\sigma}_1^2 = \frac{1}{(1-2r)}[(1-r)e_1 - r\,e_2]$$

$$\hat{\sigma}_2^2 = \frac{1}{(1-2r)}[-r\,e_1 + (r-1)e_2]$$

无论F_2或DH (BC)群体，在方差同质模型下，

$$\hat{\sigma}_j^2 = \hat{\sigma}^2 = \frac{\sum n_i e_i}{\sum n_i}$$

在 $0<r<0.5$范围内，取 r 值代入即可得到 μ_j 和 σ_j^2的估值，将这些估值代入(3)式，即得到与 r 值相对应的一系列对数似然函数值。 其中使方程(3)值最大的 r 值就是 r 的矩估计值(Luo和Wolliams 1993)，同时也就获得了QTL 基因型内数量性状均值 μ_j 和方差 σ_j^2 的估计值。

7.1.2 最大似然估计法

要获得方程(3)的最大似然解，就是要寻求未知参数 $\Phi=(r,\ \mu_j,\ \sigma_j^2)$的值，使方程(3)的值达到最大。为此，就必须就未知参数向量 Φ 对方程(3) 求偏导数，并令其等于零，即

$$\frac{\partial L}{\partial \Phi} = \sum_{i=1}^{} \sum_{k=1}^{n_i} \left[\frac{1}{f_{Mi}(x)} \sum_{j=1} \frac{\partial\, p_{ij}\ \ f_j(x)}{\partial \Phi}\right]$$

$$= \sum_{i=1}^{} \sum_{k=1}^{n_i} \left[\frac{p_{ij} f_j(x)}{f_{Mi}(x)} \sum_{j=1} \frac{\partial \ln\ f_j(x)}{\partial\ \Phi}\right]$$

若将由重组率 r 所确定的基因座位上QTL 基因型为 Q_j (F_2: j=1,2,3; DH(BC): j=1,2)的个体出现的频率定义为:

$$a(x_{ik} \mid M_i;\ Q_j) = \frac{p_{ij} f_j(x)}{f_{Mi}(x)} \tag{4}$$

则

$$\frac{\partial L}{\partial \mu_j} = \sum_{i=1} \sum_{k=1}^{n_i} \left[\frac{1}{f_{Mi}(x)} \sum_{j=1} \frac{\partial p_{ij} f_j(x)}{\partial \mu_j} \right]$$

$$= \sum_{i=1} \sum_{k=1}^{n_i} \left[\sum_{j=1} a(x_{ik} \mid M_i; Q_j) \frac{\partial \ln f_j(x)}{\partial \mu_j} \right]$$

而

$$\ln f_j(x) = -\frac{1}{2} \ln (2\pi) - \frac{1}{2} \ln \sigma_j^2 + \frac{(x-\mu_j)^2}{-\sigma_j^2}$$

$$\frac{\partial \ln f_j(x)}{\partial \mu_L} = \frac{(x-\mu_j)^2}{\sigma_j^2} \qquad (j = L)$$

$$\frac{\partial \ln f_j(x)}{\partial \mu_L} = 0 \qquad (j \neq L)$$

代入上式得:

$$\frac{\partial \ln L}{\partial \mu_j} = \sum_{i=1} \sum_{k=1}^{n_i} \left[a(x_{ik} \mid M_i; Q_j) \frac{(x_{ik}-\mu_j)^2}{\sigma_j^2} \right] \tag{5}$$

$$\frac{\partial \ln L}{\partial \sigma_j^2} = \sum_{i=1} \sum_{k=1}^{n_i} \left[\frac{1}{f_{Mi}(x)} \sum_{j} \frac{\partial p_{ij} f_j(x)}{\partial \sigma_j^2} \right]$$

$$= \sum_{j=1} \sum_{k=1}^{n_i} \left[\sum_{j=1} a(x_{ik} \mid M_i; Q_j) \frac{\partial \ln f_j(x)}{\partial \sigma_j^2} \right]$$

而

$$\frac{\partial \ln f_j(x)}{\partial \sigma_L^2} = -\frac{1}{2\sigma_j^2} + \frac{(x-\mu_j)^2}{2(\sigma_j^2)^2} \qquad (j = L)$$

$$\frac{\partial \ln f_j(x)}{\partial \sigma_L^2} = 0 \qquad (j \neq L)$$

代入上式得:

$$\frac{\partial \ln L}{\partial \sigma_j^2} = \sum_{i=1} \sum_{k=1}^{n_i} \left(a(x_{ik} \mid M_i; Q_j) \left[\frac{-1}{2\sigma_j^2} + \frac{(x_{ik}-\mu_j)^2}{2(\sigma_j^2)^2} \right] \right) \tag{6}$$

分别令方程(5)和(6)等于0，解之即得除r以外的其它未知参数的解:

$$\hat{\mu}_j = \frac{\sum_{i=1} \sum_{k=1}^{n_i} \left[a(x_{ik} \mid M_i; Q_j) x_{ik} \right]}{\sum_{i=1} \sum_{k=1}^{n_i} a(x_{ik} \mid M_i; Q_j)} \tag{7}$$

$$\hat{\sigma}_j^2 = \frac{\sum_{i=1} \sum_{k=1}^{n_i} \left[a(x_{ik} \mid M_i; Q_j) (\mu_j - x_{ik})^2 \right]}{\sum_{i=1} \sum_{k=1}^{n_i} a(x_{ik} \mid M_i; Q_j)} \tag{8}$$

对于方差同质模型，

$$\hat{\sigma}_j^2 = \hat{\sigma}^2 = \frac{1}{\sum_i n_i} \sum_{j=1} \sum_{i=1} \sum_{k=1}^{n_i} [a(x_{ik} \mid M_i; Q_j)(\mu_j - x_{ik})^2]$$

我们采用EM 算法求解上述未知参数。对于0< r <0.5 范围内特定的 r 值，先以标记基因型内的数量性状均值和方差(同质模型时用作图群体数量性状的表型方差)作为 μ_j 和 σ_j^2 的初始值代入(4)式求出 $a(x_{ik} \mid M_i; Q_j)$，然后由(7)和(8)式求出 μ_j 和 σ_j^2 及(3)式的值，再由新的 μ_j 和 σ_j^2 值求出新的 $a(x_{ik} \mid M_i; Q_j)$。如此反复迭代，使(3)式的值不断增加，直到前后两次迭代的(3)式值之差小于某一预定的临界值(如0.0001)时为止。最后比较各个r值时的对数似然值，其中与最大值相对应的 r 值就是重组率的最大似然估计值，相应的 μ_j 和 σ_j^2 就是 QTL基因型内数量性状均值和方差的估计值。

7.1.3 似然比检验

为解决 r 值和连锁的显著性检验问题，作者引入了似然比检验统计量(徐云碧1994a)。当零假设(即不存在 QTL 效应)成立时，各 QTL 基因型之间的平均数相等，似然函数变为：

$$L = L_0(\mu_p, \sigma_p^2) = \prod_{z=1}^{n} f(y_z)$$

其中：

y_z 为群体中第 z 个体数量性状的表型值；μ_p 和 σ_p^2 为作图群体的平均数和方差，可分别估计为

$$\hat{\mu}_p = \frac{1}{n} \sum_{z=1}^{n} y_z \text{ 为作图群体的平均数}$$

$$\hat{\sigma}_p^2 = \frac{1}{n} \sum_{z=1}^{n} (y_z - \mu_p)^2 \text{ 为作图群体的方差}$$

$$f(y_z) = \frac{1}{\sqrt{2\pi\sigma_p^2}} \exp\left[-\frac{(y_z - \mu_p)^2}{2\sigma_p^2}\right]$$

即平均值为 μ_p，方差为 σ_p^2 的正态概率密度函数。备择假设（标记与QTL相连锁）的似然比检验统计量可转换成 LOD 值(LOD=log of odds)

$$LOD = \log_{10}\left[\frac{L(\mu_j, \sigma_j^2)}{L_0(\mu_p, \sigma_p^2)}\right]$$

$$= \log_{10}\left[\frac{\text{Prob(标记与QTL连锁)}}{\text{Prob(不存在QTL效应)}}\right]$$

当样本较大时，$LOD \sim 1/2(\log_{10}e) \times \chi^2$ 分布 (Lander和Botstein 1989; Knott和Haley 1992)，据此可对重组率 r 及连锁进行显著性检验。对于 F_2群体，在方差同质和异质模型下其自由度(df)分别为5和7，对于 DH (BC)群体，则分别为4和5。不同自由度下的 LOD 临界值见表7.1。

表7.1 不同显著水平下的LOD临界值

df	显著水平			
	0.05	0.01	0.005	0.001
1	0.834	1.440	1.711	2.351
2	1.301	2.000	2.302	3.000
3	1.696	2.462	2.788	3.532
4	2.061	2.884	3.227	4.010
5	2.404	3.277	3.637	4.455
6	2.734	3.650	4.028	4.876
7	3.055	4.013	4.404	5.281
8	3.368	4.362	4.769	5.673
9	3.674	4.710	5.123	6.053
10	3.976	5.040	5.470	6.425

7.1.4 QTL基因效应的估计

根据采用矩法和最大似然法所得不同QTL基因型的均值和方差的估计值，可以进一步估计QTL的基因效应，包括基因加性效应、显性效应，用这些效应的估计值可进而估计显性度、遗传率等参数。

对于F_2群体，有三种QTL基因型，Q_1Q_1、Q_1Q_2和Q_2Q_2，它们的数量性状均值分别为$\hat{\mu}_1$、$\hat{\mu}_2$和$\hat{\mu}_3$。QTL的加性效应可估计为

$$\hat{d} = (\hat{\mu}_1 - \hat{\mu}_3)/2$$

显性效应为

$$\hat{h} = \hat{\mu}_2 - (\hat{\mu}_1 + \hat{\mu}_3)/2$$

显性度为

$$\hat{D} = \hat{h}/\hat{d}$$

对于DH群体，只有两种QTL基因型Q_1Q_1和Q_2Q_2，其均值为$\hat{\mu}_1$、$\hat{\mu}_2$，故只能估计基因加性效应：

$$\hat{d} = (\hat{\mu}_1 - \hat{\mu}_2)/2$$

对于回交群体，当F_1与P_1回交时，群体中的两种QTL基因型为Q_1Q_1和Q_1Q_2，它们的数量性状均值分别为$\hat{\mu}_1$和$\hat{\mu}_2$；当F_1与P_2回交时，群体中的两种QTL基因型为Q_1Q_2和Q_2Q_2，它们的数量性状均值分别为$\hat{\mu}_1'$和$\hat{\mu}_2'$。基因效应的估计要同时利用这两种回交世代，加性效应可估计为

$$\hat{d} = (\hat{\mu}_1 - \hat{\mu}_2')/2$$

显性效应可估为

$$\hat{h} = \hat{\mu}_2 - (\hat{\mu}_1 + \hat{\mu}_2')/2$$

或

$$\hat{k} = \hat{\mu}_1 - (\hat{\mu}_2 + \hat{\mu}_3)/2$$

显性度可估计为

$$\hat{D} = \hat{k}/\hat{d}$$

7.1.5 程序设计

采用Quick Basic 语言进行程序设计，编写了适应于F_2、DH和BC群体的QTL分析的通用计算程序，有关计算可在一般 PC386微机上完成（徐云碧等1994c）。

在重组率理论值范围(0～0.5)内每隔$z \times 10^{-2}$取各种可能的值代入似然方程(6)，使似然方程值最大r值即是控制数量性状的QTL与分子标记之间的重组率估计值($\hat{r}$)。对于矩估计法，在r值达到0.5之前，似然方程中的方差估值可能会出现负值。一旦出现负的方差估值，在理论上就意味着似然函数值已达最大，即可停止对更大r值作进一步分析。

7.2 Monte Carlo 模拟

Monte Carlo方法亦称为随机模拟方法或随机抽样技术。它的基本思想是，为了求解数学、物理、工程技术以及生产管理等方面的问题，首先建立一个概率模型或随机过程，使它的参数等于问题的解；然后通过对模型或过程的观察或抽样试验来计算所求参数的统计特征，最后给出所求解的近似值，而解的精确度可用估计值的标准误来表示。有关这一方法可详见 Manly (1991)。

徐云碧(1994a) 采用Monte Carlo模拟方法，以F_2群体为例模拟了矩估计法和最大似然法在 QTL 作图中的应，并比较了方差同质模型和方差异质模型。 在模拟过程中，首先由随机抽样产生标记基因型个体，使群体中三种标记基因型的比例符合共显性基因的分离比 1:2:1。 再进行第二次随机抽样产生 QTL 基因型个体，使每种标记基因型亚群体的 QTL 基因型的比例符合关系式p_{ij}(i,j =1,2,3)。这样就产生了F_2群体中每个个体的基因型。然后通过随机抽样产生满足具有 QTL 基因型均值和方差的正态分布的随机数序列，作为群体中各个体的数量性状表型值(假定仅由与特定标记相连锁的 QTL 所决定)。共模拟了4种不同的重组率(r =0, 0.15, 0.30, 0.50)，分别表示标记与 QTL 完全连锁、紧密连锁、松散连锁、完全独立。各 QTL 基因型Q_1Q_1、Q_1Q_2和Q_2Q_2的数量性状均值分别为3、6和6，方差分别为0.25、1.0和0.81。 每一种模拟组合各重复30次。F_2模拟群体大小为 n=300。

模拟结果 (表7.2) 表明，当期望重组率为0 时，不同方法和模型所得估计值均接近于0，但随着期望重组率的增加，重组率一般被低估。 就不同的方差模型而言，方差同质模型下所得重组率估计值的偏倚成度小于方差异质模型；最大似然法所估计的重组率偏倚小于矩估计法。 无论何种模型或方法，所估计的LOD 值均随重组率期望值的提高而急剧下降，其中以矩估计值更为明显。

表7.2 在方差同质和异质模型下由矩估计法 (ME)和最大似然法 (MLE)所估计的标记-QTL间重组率及其相应的 *LOD* 值

重组率		方差同质模型		方差异质模型	
(r)		MLE	ME	MLE	ME
0.00	r	0.002±0.005*	0.000±0.000	0.000±0.002	0.000±0.000
	LOD	74.524±6.568	74.490±6.580	82.901±6.027	82.895±6.027
0.15	r	0.161±0.040	0.143±0.036	0.144±0.031	0.134±0.034*
	LOD	38.455±5.121	37.007±5.334	44.313±5.184	40.970±7.199
0.30	r	0.318±0.036*	0.266±0.096	0.330±0.056**	0.230±0.079***
	LOD	22.851±3.918	18.763±6.275	25.921±4.553	16.321±8.154
0.50	r	0.369±0.182***	0.310±0.124***	0.465±0.056**	0.136±0.163***
	LOD	9.476±4.774	5.373±3.957	16.038±5.774	4.229±4.968

*P<0.05,**P<0.01, *** P<0.001

在方差异质模型下，QTL基因型内数量性状均值和方差的模拟值与其期望值符合得很好，两者之间一般不存在显著差异，只有其中4个方差模拟值显著偏离了期望值（表7.3)。QTL 基因型内数量性状均值和方差的估计结果表明，当重组率较低时，两种方法所得估计值与其对应的模拟理论值（*SIM*）符合得很好，但随着重组率的增加估计值逐渐偏离对应的模拟值，且当 r=0.50时，三种QTL基因型间的均值、方差几乎相等，这与不存在QTL效应的预期结果相吻合。就两种估计方法而言，矩估计值的偏倚程度明显高于最大似然估计值。

对于方差同质模型和异质模型，最大似然法都是一种可靠的重组率估计方法，只是在重组率过高时其估值才略为偏低。在所模拟参数下，矩估计法对紧密连锁的QTL，其重组率的估计是较为可靠的，但当连锁松驰时将导致重组率的严重低估。由此可见，参数估值的偏倚通常与松驰的标记-QTL 连锁有关。因此，在发展统计方法的同时，尽可能地饱和用于 QTL作图的分子连锁图谱也是极为重要的。

Luo和Kearsey (1989,1991)、Luo和 Woolliams (1993) 提出的以单一标记为基础的矩量法和最大似然法可以确定 QTL 与标记座位之间的重组率，但难以获得 r 估计值的抽样方差，因而难以对其进行显著性测验。我们的研究表明，在 ZYQ/JX F_2 群体中，90%以上的 r 估计值都在0.35以下，故仅根据 r 估计值的大小不能判断数量性状与特定标记间是否存在真实的关联。根据人类连锁遗传分析中的似然比概念，对不同方法和分析模型所得 *LOD* 值进行相关分析，发现存在极显著的相关(徐云碧等1995b)。*LOD* 值不仅可以定量评价不同性状与标记之间关联程度的相对大小，并能作为重组率估计值是否显著的间接标准。

表7.3 在方差异质模型下QTL基因型的数量性状均值和方差的模拟值(SIM)及其相应的矩估计值(MS)和最大似然值(MLE)

估计参数	重组率	估计方法	QTL基因型		
			Q_1Q_1	Q_1Q_2	Q_2Q_2
均值	0.00	SIM	3.002±0.057	5.980±0.094	5.991±0.120
		MLE	3.002±0.515	5.982±1.006	5.991±0.924
		ME	3.002±0.519	5.980±1.011	5.991±0.930
	0.15	SIM	2.996±0.065	6.013±0.096	5.987±0.106
		MLE	3.003±0.498	6.022±0.968	5.978±0.954
		ME	3.059±0.647	5.979±1.029	5.987±0.951
	0.30	SIM	2.982±0.063	5.985±0.075	5.996±0.102
		MLE	2.999±0.521	5.946±1.001	6.084±0.787
		ME	3.557±1.288*	5.758±1.168	5.918±0.989
	0.50	SIM	3.003±0.066	5.992±0.069	6.007±0.084
		MLE	5.109±0.933***	5.338±1.096**	4.926±0.828***
		ME	5.091±1.475***	5.343±1.435*	5.117±1.423**
方差	0.00	SIM	0.270±0.043*	1.022±0.146	0.866±0.146*
		MLE	0.266±0.043	1.011±0.140	0.853±0.144
		ME	0.270±0.043	1.022±0.146	0.866±0.146
	0.15	SIM	0.244±0.041	1.006±0.132	0.808±0.143
		MLE	0.248±0.049	0.936±0.188	0.910±0.232*
		ME	0.419±0.406**	1.058±0.367	0.904±0.346
	0.30	SIM	0.258±0.048	1.044±0.096*	0.755±0.135*
		MLE	0.271±0.064	1.001±0.265	0.619±0.340
		ME	1.659±1.238***	1.363±0.738*	0.977±0.546*
	0.50	SIM	0.256±0.029	1.025±0.122	0.825±0.156
		MLE	0.870±0.768***	1.201±0.696	0.685±0.731
		ME	2.175±0.725***	2.059±0.821***	2.024±0.931***

*P<0.05, **P<0.01, ***P<0.001

采用以单一标记为基础的最大似然法，结合 *LOD* 值的 χ^2 检验，可以有效地进行标记-QTL 关联(连锁检验)的检测及重组率的估计。与建立在两个相邻标记基础上的QTL区间作图法（详见第8章）相比，它还具有方法较为简单、运算耗时少等特点。并且，对于

只有少数可供利用的遗传标记而未建立起完整连锁图谱的物种，也可进行分析。借助于高度饱和的连锁图谱，还可对有关QTL作精确定位。其不足之处在于它每次只能利用单个标记所提供的信息。

目前在QTL 作图领域，一个具有普遍性的问题是对于标记-QTL连锁缺乏有效的统计检验方法。在具体的实验中，即使标记和 QTL 之间不存在连锁，它们之间重组率的估计值通常也小于0.50（徐云碧等1995b)。

7.3 应用与实例分析

7.3.1 应用

郑康乐等 (1992) 利用水稻三交组合和质量基因的研究方法，进行了水稻广亲和基因的 RFLP 标记定位。三交组合中标记基因型的分离类似于回交群体，而反映广亲和性强弱的结实率为质量-数量性状。在方差同质模型下，作者采用 QTLMAP -II 软件对郑康乐等(1992)表1的数据进行分析，结果表明广亲和基因与第6染色体上的色素原基因 C、DNA探针RG138和RG64 的重组率 r 分别为1%、1% 和25%，连锁极为显著（徐云碧等1995c)。其中色素原基因 C 与广亲和基因 S_p 的连锁关系已得到许多经典遗传研究的证实。而DNA探针 RG138 与广亲和基因的连锁关系进一步肯定了两个独立的遗传研究结果(刘蔼民等1992，郑康乐等1992)。这表明对于结实率之类兼有质量和数量遗传特征的性状(或称为质量-数量性状，莫惠栋 1993)，可采用以单一标记为基础的最大似然法及 QTL 作图软件 QTLMAP-II 进行有效的分析。

作者利用 ZYQ /JX F_2 群体，采用最大似然法和异质方差模型估计了与水稻产量构成性状有关的 QTL 与分布于11 条染色体上的 51 个 RFLP 标记之间的重组率，发现了7个存在显著连锁关系的标记-性状组合，其平均重组率为10.0%（徐云碧等 1995b)。这些标记分布于第 1、5、6、8 和 11 等 5 条染色体上，涉及 7个 RFLP 标记和 3 个产量构成性状，即每穗颖花数 (RG573、RZ617、RG103)、单株穗数 (RG64B) 和每穗实粒数 (RG101、RG244、RG303)。

7.3.2 实例分析

我们以第6章所给水稻 ZYQ /JX F_2群体第5染色体上的 7个 RFLP 标记为例，对生育期等5个数量性状采用矩估计法和最大似然法进行 QTL 作图。具体计算以第三个标记(RZ70)与生育期(到达抽穗的天数)为例。

1. 矩估计法

以下说明采用矩量法(方差异质模型)进行标记-QTL之间重组率估计的过程。

对于生育期-ZR70 这个性状-标记组合，F_2 群体中三种标记基因型 (ZYQ型、F_1和JX型）的个体数分别为 $n_1=9$，$n_2=41$ 和 $n_3=26$。三种标记基因型亚群体生育期的均值和方差分别为

$m_1 = 113.3333$　　$s_1^2 = 45$

$m_2 = 112.3171$　　$s_2^2 = 40.37195$

$m_3 = 110.8462$ $s_3^2 = 87.65539$

在 $0< r <0.5$ 范围内，取 r 值代入即可得到 μ_j 和 σ_j^2 的估值，将这些估值代入(3)式，即得到与 r 值相对应的一系列对数似然函数值。其中使方程(3)值最大的 r 值就是 r 的矩估计值。本例 r 的增量为0.01。

当 r =0.00 时，F_2 群体 9种标记/QTL基因型频率的矩阵为:

r(1,1)= 1 r(1,2)= 0 r(1,3)= 0

r(2,1)= 0 r(2,2)= 1 r(2,3)= 0

r(3,1)= 0 r(3,2)= 0 r(3,3)= 1

将 r 值代入方程组 (1)和(2)的解，即得QTL 三种基因型的数量性状均值的估计值:

$$\hat{\mu}_1 = \frac{1}{(1-2r)^2}\{(1-r)^2 m_1 - 2r(1-r)m_2 + r^2 m_3\}$$

$$= [(1-0)^2 \times 113.333 - 2\times 0(1-0)\times 112.3171 + 0^2 \times 110.8462]/(1-2\times 0)^2$$

$$= 113.3333$$

类似地有

$\hat{\mu}_2$=112.3171

$\hat{\mu}_3$=110.8462

将 $\hat{\mu}_1$、$\hat{\mu}_2$和 $\hat{\mu}_3$代入依次求出 d_{ij} :

$d_{ij}=(\mu_j - m_i)^2$

d_{11} =(113.3333-113.3333)^2= 0

d_{12} =(112.3171-113.3333)^2= 1.032796

d_{13} =(110.8462-113.3333)^2= 6.186078

d_{21} =(113.3333-112.3171)^2= 1.032796

d_{22} =(112.3171-112.3171)^2= 0

d_{23} =(110.8462-112.3171)^2= 2.163596

d_{31} =(113.3333-110.8462)^2= 6.186078

d_{32} =(112.3171-110.8462)^2= 2.163596

d_{33} =(110.8462-110.8462)^2= 0

由 $e_i = s_i^2 - \sum_{j=1} r_{ij}d_{ij}$ 得:

e_1 = 45 - (1×0 + 0×1.032796 + 0× 6.186078) = 45

e_2 = 40.37195 - (0×1.032796 + 1×0 + 0×2.163596) = 40.37195

e_3 = 87.65539 - (0×6.186078 + 0×2.163596 + 1×0) = 87.65539

进而方程组 (1)和(2)的解式得到三种 QTL 基因型数量性状的方差的估计值:

$$\hat{\sigma}_1^2 = \frac{1}{(1-2r)^2}\{(1-r)^2 e_1 - 2r(1-r)e_2 + r^2 e_3\}$$

$$= [(1-0)^2 \times 45 - 2\times 0(1-0)\times 40.37195 + 0^2 \times 87.65539]/(1-2\times 0)^2$$

$$= 45$$

类似地:

$\hat{\sigma}_2^2 = 40.37195$

$\hat{\sigma}_3^2 = 87.65539$

将所得到的 $\hat{\mu}_3$ 和 $\hat{\sigma}_3^2$ 与F_2群体各个体的数量性状表型值一并代入以下对数似然函数即可得到相应的似然函数值，为了简化计算，可将 (3) 式中的第一项与第二项合并，先求和。为了求LOD 值的方便，取以10为底的对数，即：

$$\log_{10} L = \sum_{i=1}^{3} \sum_{k=1}^{n_i} \log_{10} \{ \sum_{j=1} \frac{p_{ij}}{\sqrt{2\pi}\,\sigma_j} \exp [- \frac{(x_{ik} - \mu_j)^2}{2\sigma_j^2}] \}$$

其中，$i = 1$ 时，$k = 9$；$i = 2$ 时，$k = 41$；$i = 3$时，$k = 26$。

我们先考虑对数符号内的求和 SUM。

$$SUM = \sum_{j=1} \frac{p_{ij}}{\sqrt{2\pi}\,\sigma_j} \exp [- \frac{(x_{ik} - \mu_j)^2}{2\sigma_j^2}] \}$$

例如当 $i = 1$，$k = 1$时，x_{ik} 为112，则上述求和的值 SUM 为

$$SUM = \frac{1}{\sqrt{2\times 3.1415926\times 45}} \exp [- \frac{(112 - 113.3333)^2}{2\times 45}]$$
$$+ \frac{0}{\sqrt{2\times 3.1415926\times 40.37195}} \exp [- \frac{(112 - 112.3171)^2}{2\times 40.37195}]$$
$$+ \frac{0}{\sqrt{2\times 3.1415926\times 87.65539}} \exp [- \frac{(112 - 110.8462)^2}{2\times 87.65539}]$$
$$= 0.0583076$$

对于不同的 i 和 k，对应的 x_{ik} 和 SUM 如表7.4。

对 SUM 值取对数，并对 i和 k 求和，即得似然函数的对数值：

$$\sum_{i=1}^{3} \sum_{k=1}^{n_i} \log_{10} SUM_{ik}$$
$$= [\log_{10}(0.0583076) + \log_{10}(5.766333E-02) + \ldots + \log_{10}(5.121861E-02)]$$
$$+ [\log_{10}(4.423934E-02) + \log_{10}(1.285262E-02) + \ldots + \log_{10}(6.062286E-02)]$$
$$+ [\log_{10}(1.149846E-02) + \log_{10}(1.017357E-02) + \ldots + \log_{10}(4.228853E-02)]$$
$$= [-12.76844] + [-57.98129] + [-41.05105]$$
$$= -111.8028$$

LOD值为标记-QTL连锁和不存在与该标记有关的QTL 效应两种可能性之比取对数，即：

$$LOD = \log_{10} \frac{L(\mu_j, \sigma_j^2)}{L_0(\mu_p, \sigma_p^2)}$$
$$= \log_{10} L(\mu_j, \sigma_j^2) - \log_{10} L_0(\mu_p, \sigma_p^2)$$

其中

$$\log_{10} L_0(\mu_p, \sigma_p^2) = \sum_{z=1}^{n} \log_{10} f(y_z) = \sum_{z=1}^{n} \log_{10} \frac{1}{\sqrt{2\pi}\,\sigma_p^2} \exp [- \frac{(y_z - \mu_p)2}{2\sigma_p^2}]$$

将作图群体数量性状均值 $\hat{\mu}_p$ (=111.7342) 和方差 $\hat{\sigma}_p^2$ (=55.53514) 以及群体中第 z 个体的数量性状表型值(见表6.11) 代入即得：

表 7.4　矩估计法估计标记-QTL重组率时与似然函数有关的计算结果

i	k	x_{ik}	SUM	i	k	x_{ik}	SUM	i	k	x_{ik}	SUM
1	1	112	0.0583076	2	17	114	6.062286E-02	3	1	126	1.149846E-02
	2	115	5.766333E-02		18	114	6.062286E-02		2	95	1.017357E-02
	3	110	5.256385E-02		19	119	3.611192E-02		3	118	3.182288E-02
	4	124	1.679848E-02		20	105	3.235184E-02		4	115	3.861683E-02
	5	102	1.427235E-02		21	113	6.242544E-02		5	104	3.261452E-02
	6	106	3.271895E-02		22	119	3.611192E-02		6	109	4.179049E-02
	7	119	4.162478E-02		23	116	5.307795E-02		7	123	0.0183479
	8	115	5.766333E-02		24	114	6.062286E-02		8	109	4.179049E-02
	9	117	5.121861E-02		25	109	5.478845E-02		9	107	3.916287E-02
					26	106	3.830298E-02		10	124	1.588152E-02
2	1	107	4.423934E-02		27	113	6.242544E-02		11	106	3.726844E-02
	2	101	1.285262E-02		28	108	4.984568E-02		12	92	5.618824E-03
	3	119	3.611192E-02		29	125	8.56398E-03		13	115	3.861683E-02
	4	107	4.423934E-02		30	120	3.022605E-02		14	121	2.366526E-02
	5	114	6.062286E-02		31	101	1.285262E-02		15	122	2.095686E-02
	6	111	6.145256E-02		32	124	1.158061E-02		16	112	4.228853E-02
	7	119	3.611192E-02		33	124	1.158061E-02		17	114	4.026056E-02
	8	113	6.242544E-02		34	112	6.270894E-02		18	106	3.726844E-02
	9	116	5.307795E-02		35	114	6.062286E-02		19	89	2.80044E-03
	10	113	6.242544E-02		36	118	0.0420884		20	117	3.433258E-02
	11	115	0.057432		37	108	4.984568E-02		21	114	4.026056E-02
	12	101	1.285262E-02		38	104	2.665681E-02		22	111	4.260515E-02
	13	109	5.478845E-02		39	99	6.98223E-03		23	109	4.179049E-02
	14	111	6.145256E-02		40	111	6.145256E-02		24	103	2.999259E-02
	15	110	5.874801E-02		41	114	6.062286E-02		25	109	4.179049E-02
	16	115	0.057432						26	112	4.228853E-02

$$\log_{10} L_o(\mu_p, \sigma_p^2)$$

$$= \log_{10} \frac{\exp\left[-\frac{(126-111.7342)^2}{2\times 55.53514}\right]}{\sqrt{2\times 3.1415926\times 55.53514}} + \log_{10} \frac{\exp\left[-\frac{(107-111.7342)^2}{2\times 55.53514}\right]}{\sqrt{2\times 3.1415926\times 55.53514}}$$

$$+ \ \ldots \ + \log_{10} \frac{\exp\left[-\frac{(117-111.7342)^2}{2\times 55.53514}\right]}{\sqrt{2\times 3.1415926\times 55.53514}}$$

$$= -113.1276$$

因此，当 r = 0.00时，　LOD = -111.8028-(-113.1276)　= 1.3248

当 r =0.01 时，F_2 群体 9种标记/QTL基因型频率的矩阵为：

r(1,1)= 0.9801　　r(1,2)= 0.0198　　r(1,3)= 0.0001

r(2,1)= 0.0099　　r(2,2)= 0.9802　　r(3,3)= 0.0099

r(3,1)= 0.0001　　r(3,2)= 0.0198　　r(3,3)= 0.9801

三种 QTL 基因型数量性状均值的估计值:

$\hat{\mu}_1$= 113.354

$\hat{\mu}_2$= 112.3218

$\hat{\mu}_3$= 110.8161

进而求出各 d_{ij} 及 e_i:

d_{11} = 4.281141E-04　d_{12} = 1.023297　　d_{13} = 6.336548

d_{21} = 1.075279　　d_{22} = 2.194406E-05　d_{23} = 2.252954

d_{31} = 6.28943　　d_{32} = 2.177399　　d_{33} = 9.040512E-04

e_1 = 44.97869　　e_2 = 40.33898　　e_3 = 87.61076

三种 QTL 基因型数量性状方差的估计值为:

$\hat{\sigma}_1^2$ = 45.07878

$\hat{\sigma}_2^2$ = 39.80387

$\hat{\sigma}_3^2$ = 88.5809

类似于 r =0.00时的计算步骤，将 $\hat{\mu}_i$ 和 $\hat{\sigma}_i^2$ 与F_2群体各个体的数量性状表型值一并代入似然函数获得相应的对数似然函数值，其中三种标记基因型亚群体 SUM值的对数和分别为-12.76862，-57.98129 和-41.05105，即得

$$\sum_{j=1}^{3}\sum_{k=1}^{n_j}\log_{10} SUM_{jk} = (-12.76862) + (-57.98129) + (-41.05105) = -111.801$$

表 7.5　与 r 值相对应的 LOD 估计值

r	LOD	r	LOD
0.00	1.3248	0.13	1.19381
0.01	1.3266	0.14	1.146889
0.02	1.327782	0.15	1.086304
0.03	1.328072	0.16	1.008263
0.04	1.327271	0.17	0.9078293
0.05	1.32515	0.18	0.7786942
0.06	1.321335	0.19	0.6120377
0.07	1.315422	0.20	0.3932419
0.08	1.306877	0.21	0.0880508
0.09	1.295044	0.22	-0.407402
0.10	1.279053	0.23	-1.412773
0.11	1.257835	0.24	-3.808327
0.12	1.230026		

由于 $\log_{10} L_0(\mu_p, \sigma_p^2)$ 与重组率无关，即等于-113.1276。因此，当 r=0.01时，

LOD =-111.801-(-113.1276) = 1.3266

采用类似的方法，可求出与不同 r 值对应的LOD 值如表7.5。

在本例中，当 r =0.25时，三种 QTL 基因型数量性状方差出现负估值，此时停止对更大的 r 作进一步分析（即认为在此值之前，似然函数值已达最大）。

由上表可见，当 r =0.03时，LOD 值最大（为1.328072）。此时三种 QTL 基因型的数量性状均值和方差分别为

$\hat{\mu}_1$= 113.398　　$\hat{\sigma}_1^2$ = 45.28241

$\hat{\mu}_2$= 112.332　　$\hat{\sigma}_2^2$ = 38.55648

$\hat{\mu}_3$= 110.752　　$\hat{\sigma}_3^2$ = 90.58125

由其中的均值估计值进一步估计出 QTL的加性效应、显性效应和显性度分别为

$$\hat{d} = (\hat{\mu}_1 - \hat{\mu}_3)/2$$

$$= (113.398 - 110.752)/2 = 1.323$$

$$\hat{h} = \hat{\mu}_2 - (\hat{\mu}_1 + \hat{\mu}_3)/2$$

$$= 112.332 - (113.398 + 110.752)/2 = 0.257$$

$$\hat{D} = \hat{h}/\hat{d}$$

$$= 0.257 / 1.323 = 0.194$$

采用以上分析方法，可对所有性状/标记组合进行分析。表7.6列出了方差同质模型和方差异质模型下，标记-QTL 之间重组率的估计值及相应的 LOD 值，表7.7给出了QTL 基因的加性效应、显性效应和显性度的估计值。

2. 最大似然法

最大似然法是采用EM 算法求解未知参数。本例考虑方差异质模型。对于 $0<r<0.5$ 范围内特定的 r 值，先以三种标记基因型亚群体内数量性状的均值（113.3333，112.3171和110.1276）和方差（45，40.37195 和87.65539）作为 μ_j 和 σ_j^2 的初始值代入(4)式求出 $a(x_{ik} \mid M_i; Q_j)$，然后由(7)和(8)式求出 μ_j 和 σ_j^2 及(3)式的值，再由新的 μ_j 和 σ_j^2 值求出新的 $a(x_{ik} \mid M_i; Q_j)$。如此反复迭代，使(3)式的值不断增加，直到前后两次迭代的(3)式值之差小于某一预定的临界值（本例为0.0001）时为止。

(4) 式是由 标记-QTL之间重组率为 r 的基因座位上QTL 基因型为 Q_j 的个体出现的频率，其中 p_{ij} 与矩估计法相同。例如，当 r=0.00 时，F_2群体9种标记/QTL基因型频率的矩阵为：

r(1,1)= 1　　r(1,2)= 0　　r(1,3)= 0

r(2,1)= 0　　r(2,2)= 1　　r(2,3)= 0

r(3,1)= 0　　r(3,2)= 0　　r(3,3)= 1

当 r =0.01 时，上述矩阵变为：

r(1,1)= 0.9801　　r(1,2)= 0.0198　　r(1,3)= 0.0001

r(2,1)= 0.0099　　r(2,2)= 0.9802　　r(2,3)= 0.0099

r(3,1)= 0.0001　　r(3,2)= 0.0198　　r(3,3)= 0.9801

以下根据(4)式求 $a(x_{ik} \mid M_i; Q_j)$。对于特定的 r 值，先以三种标记基因型亚群体内数量性状的均值(113.3333, 112.3171和110.1276)和方差(45,40.37195和87.65539)作为 μ_j 和 σ_j^2 的初始值，连同各个 x_{ik} 一并代入(4)式求出 $a(x_{ik} \mid M_i; Q_j)$，结果列

表7.6 ZYQ/JX F_2群体标记-QTL之间重组率的估计及相应的LOD值

标记	生育期		株高		穗长		颖花数		结实率	
	r	LOD	r	LOD	r	LOD	r	LOD	r	LOD
				矩估计值	(方差同质模型)					
RG573	0.00	0.239	0.00	1.070	0.00	1.764	0.11	2.145	0.29	0.384
RG13	0.10	0.612	0.00	0.348	0.00	1.397	0.00	0.473	0.00	0.008
RZ70	0.37	0.478	0.00	1.601	0.00	2.204	0.00	2.839*	0.37	1.897
RG697	0.00	0.286	0.05	0.718	0.23	0.069	0.00	0.092	0.15	3.350**
RG474	0.00	0.838	0.00	0.723	0.12	0.345	0.07	0.945	0.00	0.078
RG435	0.18	0.400	0.00	0.431	0.00	0.075	0.00	0.119	0.31	1.136
RZ225	0.00	0.011	0.00	1.507	0.00	0.404	0.07	0.078	0.24	1.118
				矩估计值	(方差异质模型)					
RG573	0.00	0.294	0.00	1.295	0.00	2.914	0.06	2.064	0.00	0.414
RG13	0.00	0.873	0.00	0.938	0.00	1.708	0.00	1.898	0.00	0.131
RZ70	0.03	1.328	0.00	2.319	0.00	2.785	0.08	4.839**	0.00	0.674
RG697	0.22	0.730	0.00	0.792	0.00	0.329	0.25	1.644	0.00	2.742
RG474	0.18	1.099	0.00	0.759	0.00	1.095	0.07	1.268	0.00	0.324
RG435	0.00	1.404	0.00	0.491	0.00	0.387	0.00	0.211	0.00	1.245
RZ225	0.05	0.358	0.00	1.711	0.00	0.797	0.27	1.439	0.00	0.872
				最大似然估计值	(方差同质模型)					
RG573	0.44	0.426	0.00	1.083	0.16	1.991	0.10	2.165	0.34	1.072
RG13	0.10	0.631	0.36	0.489	0.00	1.413	0.21	0.498	0.00	0.023
RZ70	0.38	0.543	0.00	1.612	0.29	2.924*	0.26	6.031***	0.39	2.101
RG697	0.48	0.336	0.35	0.813	0.40	0.622	0.05	0.105	0.15	3.549**
RG474	0.33	0.952	0.00	0.737	0.05	0.351	0.13	0.969	0.31	0.817
RG435	0.19	0.506	0.40	0.468	0.35	0.182	0.00	0.133	0.36	2.650*
RZ225	0.47	0.354	0.00	1.519	0.00	0.420	0.31	0.143	0.24	1.249
				最大似然估计值	(方差异质模型)					
RG573	0.43	1.170	0.00	1.310	0.00	2.928	0.14	4.463**	0.39	2.766
RG13	0.41	1.194	0.36	1.344	0.00	1.723	0.42	3.529*	0.42	2.592
RZ70	0.06	1.352	0.00	2.339	0.29	3.039	0.27	6.261***	0.41	2.606
RG697	0.43	1.414	0.37	0.912	0.30	0.698	0.47	2.995	0.14	3.774*
RG474	0.35	1.766	0.00	0.776	0.00	1.111	0.21	3.622*	0.44	1.201
RG435	0.41	1.472	0.28	0.511	0.31	0.690	0.44	2.692	0.38	3.011
RZ225	0.46	0.750	0.00	1.726	0.00	0.812	0.30	2.364	0.24	1.309

* $P<0.05$, ** $P<0.01$, *** $P<0.001$

于表 7.8。

将表7.8 的结果代入(7)和(8)式可分别求出新的 μ_j 和 σ_j^2 值。当 r =0时，每轮迭代所得新值与初始值完全一样，即 μ_j 的估计值分别为113.3333,112.3171和110.8462，σ_j^2 的估计值分别为 45，40.437195和87.65539。将 μ_j 和 σ_j^2 估计值代入(3)式求对数似然值进而求 LOD 值的过程完全类似于矩估计法。由此经前后两次迭代所得对数似然值相等(均为 -111.7828) 。故当 r =0时，迭代两次即停止。

表7.7 ZYQ/JX F_2群体 QTL 基因加性效应 d、显性效应 h 和显性度(D)的估计

标记	生育期			株高			穗长			颖花数			结实率		
	d	*h*	*D*	*d*	*h*	*D*	*d*	*h*	*D*	*d*	*h*	*D*	*d*	*h*	*D*
							矩估计值 (方差同质模型)								
RG573	1.10	-1.12	1.02	3.00	3.18	1.06	0.48	1.54	3.21	2.09	71.28	34.15	4.39	-26.68	6.07
RG13	1.79	-3.76	2.10	1.81	1.58	0.87	0.55	1.21	2.19	-8.35	19.97	2.39	-0.56	-1.07	1.93
RZ70	4.78	3.36	0.70	4.22	2.45	0.58	1.38	0.38	0.27	34.47	13.54	0.39	-22.77	-2.79	0.12
RG697	0.75	1.79	2.38	3.75	1.88	0.50	0.21	-1.20	5.75	-5.46	9.53	1.75	2.14	30.96	14.46
RG474	2.55	-0.50	0.20	2.31	3.11	1.34	0.57	0.84	1.47	-1.14	41.87	36.66	2.06	0.61	0.29
RG435	2.33	-2.10	0.90	1.98	1.84	0.93	-0.00	0.40	112.40	-4.85	9.59	1.98	18.21	-8.27	0.45
RZ225	-0.39	0.33	0.85	2.99	4.39	1.47	0.36	0.62	1.71	1.70	11.61	6.84	5.27	26.32	5.00
							矩估计值 (方差异质模型)								
RG573	1.10	-1.12	1.02	3.00	3.18	1.06	0.48	1.54	3.21	1.85	56.00	30.27	1.85	-4.71	2.55
RG13	1.43	-2.41	1.68	1.81	1.58	0.87	0.55	1.21	2.19	-8.35	19.97	2.39	-0.56	-1.07	1.93
RZ70	1.32	0.25	0.19	4.22	2.45	0.58	1.38	0.38	0.27	41.03	19.19	0.47	-5.92	-0.19	0.03
RG697	1.34	5.70	4.25	3.38	1.53	0.45	0.11	-0.35	3.10	-10.92	38.10	3.49	1.50	15.17	10.12
RG474	3.98	-1.22	0.31	2.31	3.11	1.34	0.43	0.49	1.12	-1.14	41.87	36.66	2.06	0.61	0.29
RG435	1.49	-0.86	0.58	1.98	1.84	0.93	-0.00	0.40	112.40	-4.85	9.59	1.98	6.92	-1.20	0.17
RZ225	-0.43	0.41	0.94	2.99	4.39	1.47	0.36	0.62	1.71	3.17	40.57	12.79	2.74	7.12	2.60
							最大似然估计值 (方差同质模型)								
RG573	5.29	4.56	0.86	3.00	3.18	1.06	-0.28	3.74	13.38	1.23	68.88	55.95	-7.02	-29.22	4.16
RG13	1.79	-3.74	2.09	6.87	-6.93	1.01	0.55	1.21	2.19	2.32	64.00	27.61	-0.56	-1.07	1.93
RZ70	4.43	5.17	1.17	4.22	2.45	0.58	3.46	-0.29	0.08	82.26	-29.11	0.35	-24.05	-2.29	0.10
RG697	5.24	4.87	0.93	11.27	-0.01	0.00	1.54	-3.58	2.33	-6.05	11.59	1.92	3.68	30.76	8.36
RG474	5.78	3.46	0.60	2.31	3.11	1.34	0.48	0.60	1.24	-5.48	54.70	9.99	2.96	28.60	9.66
RG435	2.72	-3.68	1.35	9.83	-2.35	0.24	0.19	3.16	16.32	-4.85	9.59	1.98	24.08	-1.81	0.08
RZ225	-5.08	5.09	1.00	2.99	4.39	1.47	0.36	0.62	1.71	-5.23	57.17	10.94	2.94	26.74	9.09
							最大似然估计值 (方差异质模型)								
RG573	7.46	-0.91	0.12	3.00	3.18	1.06	0.48	1.54	3.21	3.34	81.73	24.47	-11.32	-30.68	2.71
RG13	7.89	-0.97	0.12	7.37	-7.63	1.04	0.55	1.21	2.19	-39.05	-88.11	2.26	-11.57	-29.87	2.58
RZ70	1.43	0.37	0.26	4.22	2.45	0.58	3.35	-0.17	0.05	81.82	-25.21	0.31	-25.55	-2.71	0.11
RG697	8.29	-0.67	0.08	11.00	0.54	0.05	0.50	-2.95	5.94	-22.87	75.07	3.28	4.34	29.16	6.73
RG474	7.85	-0.60	0.08	2.31	3.11	1.34	0.43	0.49	1.12	-11.30	79.10	7.00	9.17	29.46	3.21
RG435	7.94	-0.86	0.11	4.52	6.27	1.39	-0.23	3.00	13.24	-24.65	71.35	2.89	24.94	-2.39	0.10
RZ225	-3.48	3.98	1.14	2.99	4.39	1.47	0.36	0.62	1.71	-8.83	71.15	8.05	3.91	25.35	6.49

相应的LOD值显著者用下划线标出

表7.8　最大似然法估计标记-QTL重组率时与似然函数有关的计算结果(r=0.00)

i	k	x_{ik})	$f_{Mi}(x)$	$a(x_{ik} \mid M_i; Q_j)$		
				j=1	j=2	j=3
1	1	112	0.0583076	1	0	0
	2	115	5.766333E-02	1	0	0
	3	110	5.256385E-02	1	0	0
	...		...		...	
	9	117	5.121861E-02	1	0	0
2	1	107	4.423934E-02	0	1	0
	2	101	1.285262E-02	0	1	0
	3	119	3.611192E-02	0	1	0
	...		...		...	
	41	114	6.062286E-02	0	1	0
3	1	126	1.149846E-02	0	0	1
	2	95	1.017357E-02	0	0	1
	3	118	3.182288E-02	0	0	1
	...		...		...	
	26	112	4.228853E-02	0	0	1

表7.9　最大似然法估计标记-QTL重组率时与似然函数有关的计算结果(r=0.01)

i	k	$x\ ik$)	$f_{Mi}(x)$	$a(x_{ik} \mid M_i; Q_j)$		
				j =1	j =2	j =3
1	1	112	6.172569E-02	0.9795656	2.036466E-02	6.984598E-05
	2	115	6.086591E-02	0.9810627	1.887292E-02	6.4448E-05
	3	110	5.498616E-02	0.9785394	2.138184E-02	7.869361E-05
	...		...		...	
	9	117	5.321572E-02	0.9820306	1.790422E-02	6.522748E-05
2	1	107	4.429178E-02	8.539679E-03	0.9825635	8.896875E-03
	2	101	1.259408E-02	7.406037E-03	0.9733748	1.921919E-02
	3	119	0.0360525	1.159469E-02	0.9803618	8.043525E-03
	...		...		...	
	41	114	6.113456E-02	1.015818E-02	0.9832081	6.633753E-03
3	1	126	1.102387E-02	7.700954E-05	1.060182E-02	0.9893212
	2	95	9.630433E-03	9.808299E-06	2.90381E-03	0.9970863
	3	118	3.227814E-02	1.488494E-04	2.587846E-02	0.9739727
	...		...		...	
	26	112	4.351816E-02	1.417616E-04	2.888502E-02	0.9709733

当 r =0.01 时，以前一代 r 值时所得均值和方差的估计值，即分别为 113.3333，112.3171和 110.8462及40，39.38727 和84.28402作为 μ_j 和 σ_j^2 的初始值，连同各个 x_{jk} 一并代入(4)式求 $a(x_{jk} \mid M_i; Q_j)$，所得结果如表7.9。

将上述结果代入(7)和(8)式所求出新的 μ_j 和 σ_j^2 值分别为 113.352，112.3261和110.8077及 39.91246，39.11231和84.87943，相应的对数似然函数值 为-111.7808。第二轮迭代结果与第一轮完全相同，因此，当 r =0.01时，迭代两次即收敛。

对于不同的 r 值，每次迭代所得对数似然函数值列于表7.10。

表7.10 不同 r 值下各迭代轮次的对数似然函数值

r	迭代轮次				
	1	2	3	4	5
0.00	-111.7828	-111.7828			
0.01	-111.7808	-111.7808			
0.02	-111.7791	-111.7792			
0.03	-111.7778	-111.7777			
0.04	-111.7767	-111.7767			
0.05	-111.7759	-111.7759			
0.06	-111.7756	-111.7755			
0.07	-111.7756	-111.7756			
0.08	-111.7761	-111.7761			
0.09	-111.777	-111.777			
0.10	-111.7784	-111.7784			
0.15	-111.7935	-111.7934			
0.20	-111.8226	-111.8225	-111.8224		
0.25	-111.8647	-111.8646	-111.8645	-111.8644	
0.30	-111.9181	-111.9178	-111.9177	-111.9176	-111.9175
0.35	-111.9834	-111.9831	-111.983	-111.9829	-111.9828
0.40	-112.0677	-112.0676	-112.0674	-112.0674	
0.45	-112.182	-112.1819	-112.1818		
0.50	-112.3342	-112.3341	-112.334		

由上表可见，当 r=0.06时，对数似然函数值达到最大(-111.7755)。此 r 值即为标记-QTL之间重组率的最大似然估计值，此时 LOD =(-111.7755)-(-113.1276)=1.3521。与该 r 值相对应的 QTL基因型数量性状均值和方差的估计值分别为113.446,112.386和110.581及 39.48546，37.68335 和 87.87891。由这些估计值进一步得到QTL的基因加性性效应、显性效应及显性度的估计值分别为1.432,0.372 和0.260。

在方差同质和异质两种模型下，所有其它标记/性状的重组率和 LOD估计值列于表7.6，

相应的加性效应、显性效应和显性度的估计值列于表7.7。

3. 结果分析

根据不同方法和模型的分析，矩估计法所检测到的QTL少于最大似然法。在第5染色体上检测到与5个性状有关的QTL数分别为2个和7个，其中极显著者分别为2个和3个。不同方法和模型均揭示出在第5染色体上存在1个影响每穗颖花数的QTL，该QTL与RZ70连锁，重组率为0%～27%。在第5染色体上还存在一个影响结实率的QTL，它与标记RG697相连锁，重组率为14%～15%。此外，还可能存在与标记RG573、RG13和RG474有关、决定颖花数的QTL（表7.6）。在第6章中所揭示的标记-QTL之间的连锁关系得到了较高LOD值的证实。值得一提的是，尽管不同方法和模型所估计的重组率和LOD值存在较大差异，但它在反映标记-QTL连锁关系的趋势上是一致的。

在基因效应的估计中，负的加性效应估计值表示来自ZYQ的QTL等位基因为减效等位基因，而来自JX的QTL等位基因为增效等位基因，负的显性效应估计值表示减效等位基因为显性，增效等位基因为隐性，反之亦然。在理论上，只有那些与标记存在显著连锁关系的QTL，其基因效应及显性度的估计值才是有意义的，因此我们只考虑表7.7中划线部分的基因效应和显性度的估计值。不同方法和模型所得估计值反映了相似的结果。对于与标记RZ70连锁、决定颖花数的QTL，来自ZYQ的等位基因具有增加颖花数的作用，单个等位基因的效应可使颖花数增加34.5～82.3个，该QTL表现为部分显性，显性度的估计值为0.31～0.47。与标记RG697相连锁、决定结实率的QTL则表现为超显性，加性效应的作用很小，单个QTL的显性效应可使结实率提高29.2～31.0个百分点（表7.7）。

第 8 章　QTL作图：区间作图法

表现连续变异的数量性状是动植物育种工作者的主要选择对象。为提高选择效率，必须在数量性状和可以作为选择指标的遗传标记之间建立起关联和连锁关系。采用经典的遗传标记，虽然可以偶尔检测到数量性状基因座位 (Quantitative trait locus, QTL)与遗传标记之间的相互关联，但由于数量有限和表型有害性，无法对整个基因组进行分析。由于RFLP等分子标记的发展，已在许多植物中构建了相对饱和的分子标记连锁图谱，因而有可能在整个基因组范围内进行 QTL定位和作图。已倡导了几种 QTL作图方法，其中包括 Weller (1986)、Luo和Kearsey (1989,1991)所发展的单个标记与QTL连锁的估计方法。但这类以单一标记为基础的作图方法，其效率将随着标记与QTL之间实际距离的增加而下降。为了提高效率，每一染色体上就必须定位更多的遗传标记。Jensen(1989)、Lander和Botstein(1989)等发展了以连锁图谱上两个相互侧连的分子标记为基础的 QTL作图方法，可以同时利用两个共显性标记的分离信息，从而获得这两个标记之间某个染色体片段上有关 QTL 的最大连锁信息。其中，Lander和Botstein 的方法最受遗传育种学家的推崇。

虽然后来van Ooijen(1992) 对区间作图法作了较为系统的描述，但由于其统计技术的复杂性以及模型倡导者简要处理问题的方式使这一方法及有关程序难以被动植物育种工作者所理解，而现有的计算机软件需要借助于大型计算机或计算机工作站才能完成其分析(Lincoln等 1993)，因而限制了该方法的有效利用。

本章将讨论QTL 区间作图方法的一些理论与应用问题，并提供应用实例。

8.1　遗传学原理与统计方法

Lander和 Botstein (1989)和 van Ooijen(1992)所介绍的区间作图法可用于 F_2和回交 (BC) 群体数量性状的基因定位，我们在此将其扩展到双单倍体 (DH)群体(徐云碧等 1995a)。

8.1.1 基本假定

设亲本P_1和P_2基因型分别为$M_{a1}M_{a1}Q_1Q_1M_{b1}M_{b1}$和$M_{a2}M_{a2}Q_2Q_2M_{b2}M_{b2}$，$F_1$自交产生 F_2；F_1 与P_2 回交产生 BC群体；F_1经花药培养或其它途径产生单倍体经加倍形成DH群体。遗传图谱中两个相邻的标记座位M_a和M_b将染色体分割成由这两个标记所包括的区间。在该区间内的QTL与标记座位M_a和M_b的图距分别为 θ_a和 θ_b，M_a和M_b之间的图距为 θ。则可采用双曲正切函数将图距 θ 转换成相应的重组率 r：

$$r=\frac{1}{2}\tanh(2\theta)=\frac{1}{2}\cdot\frac{e^{2\theta}-e^{-2\theta}}{e^{2\theta}+e^{2\theta}}$$

在理论上，$0<r_a<r$，无干扰时 $r=r_a+r_b-2r_ar_b$。

设作图群体中 n 个个体或系的数量性状表型值为 $y=\{y_1,y_2\cdots y_n\}$，假定QTL基因型 Q_1Q_1，Q_1Q_2和Q_2Q_2的效应分别服从正态分布 $N(\mu_1,\sigma^2)$，$N(\mu_2,\sigma^2)$和 $N(\mu_3,\sigma^2)$，则该 QTL 对数量性状的影响表现为三种正态分布的混合物，各分布所占比例对于标记座位 M分别为P_{m1}、P_{m2}和 P_{m3}。对于 BC群体 P_{m1} 为零，对于 DH群体 P_{m2}为零。则某一个体或品系的表型值具有以下概率密度函数：

$$f(y\mid m;r_a)=\sum_{q=1}^{}P_{mq}f_q(y) \tag{1}$$

其中，m — 标记基因型

r_a — 标记M_a与QTL之间的重组率

P_{mq} — QTL基因型的概率($q\in\{1,2,3\}$)，取决于标记基因型m和r_a

$$f_q(y)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(y-\mu_q)^2}{2\sigma^2}\right] \tag{2}$$

即平均值为 μ_q，方差为 σ^2的正态概率密度函数。

8.1.2 似然函数

对于特定的M_a和M_b，F_2和DH (BC)群体中分别有9种和4种可能的标记基因型，群体中某一个体或系应取其中之一。具有某种标记基因型的个体或系，其 QTL 基因型为 Q_1Q_1、Q_1Q_2和 Q_2Q_2 的概率之和$P_{m1}+P_{m2}+P_{m3}$为1。利用表8.1、表8.2和表8.3的期望基因型频率可计算出某一个体和品系的各类QTL基因型的概率P_{m1}、P_{m2}、P_{m3}。作图群体中全部个体或系的联合概率密度或似然函数为：

$$L=L(\mu_q,\mu_1,\mu_2,\sigma^2,y_1,y_2\cdots y_n)$$

$$=\prod_{i=1}^{n}f(y_i\mid m_i;r_a)=\prod_{i=1}^{n}\sum_{q=1}^{}P_{mq}f_q(y)$$

表 8.1　F_2群体的期望基因型频率(乘以4)

基因型	Q_1Q_1	Q_1Q_2	Q_2Q_2
$M_{a1}M_{a1}M_{b1}M_{b1}$	$r_a^2r_b^2$	$2r_a(1-r_a)\ r_b(1-r_b)$	$(1-r_a)^2(1-r_b)^2$
$M_{a1}M_{a1}M_{b1}M_{b2}$	$2r_a^2r_b(1-r_b)$	$2\ r_a(1-r_a)[r_b^2+(1-r_b)^2]$	$2(1-r_a)^2r_b(1-r_b)$
$M_{a1}M_{a1}M_{b2}M_{b2}$	$r_a^2(1-r_b)^2$	$2r_a(1-r_a)r_b(1-r_b)$	$(1-r_a)^2r_b^2$
$M_{a1}M_{a2}M_{b1}M_{b1}$	$2r_a(1-r_a)\ r_b^2$	$2[r_a^2+(1-r_a)^2r_b(1-r_b)$	$2r_a(1-r_a)(1-r_b)^2$
$M_{a1}M_{a2}M_{b1}M_{b2}$	$4r_a(1-r_a)\ r_b(1-r_b)$	$2[r_a^2+(1-r_a)^2][r_b^2+(1-r_b)^2]$	$4r_a(1-r_a)r_b(1-r_b)$
$M_{a1}M_{a2}M_{b2}M_{b2}$	$2r_a(1-r_a)(1-r_b)^2$	$2\{r_a^2+(1-r_a)^2\ r_b(1-r_b)$	$2\ r_a(1-r_a)r_b^2$
$M_{a2}M_{a2}M_{b1}M_{b1}$	$(1-r_a)^2\ r_b^2$	$2r_a(1-r_a)r_b(1-r_b)$	$r_a^2(1-r_b)^2$
$M_{a2}M_{a2}M_{b1}M_{b2}$	$2(1-r_a)^2r_b(1-r_b)$	$2r_a(1-r_a)[r_b^2+(1-r_b)^2]$	$r_a^2r_b(1-r_b)$
$M_{a2}M_{a2}M_{b2}M_{b2}$	$(1-r_a)^2(1-r_b)^2$	$2r_a(1-r_a)r_b(1-r_b)$	$r_a^2r_b^2$

表 8.2　DH 群体的期望基因型频率(乘以2)

基因型	Q_1Q_1	Q_2Q_2
$M_{a1}M_{a1}M_{b1}M_{b1}$	$r_a r_b$	$(1-r_a)(1-r_b)$
$M_{a1}M_{a1}M_{b2}M_{b2}$	$r_a(1-r_b)$	$(1-r_a)r_b$
$M_{a2}M_{a2}M_{b1}M_{b1}$	$(1-r_a)r_b$	$r_a(1-r_b)$
$M_{a2}M_{a2}M_{b2}M_{b2}$	$(1-r_a)(1-r_b)$	$r_a r_b$

表8.3　BC群体的期望基因型频率(乘以2)

基因型	Q_1Q_2	Q_2Q_2
$M_{a1}M_{a2}M_{b1}M_{b1}$	$r_a r_b$	$(1-r_a)(1-r_b)$
$M_{a1}M_{a2}M_{b2}M_{b2}$	$r_a(1-r_b)$	$(1-r_a)r_b$
$M_{a2}M_{a2}M_{b1}M_{b2}$	$(1-r_a)r_b$	$r_a(1-r_b)$
$M_{a2}M_{a2}M_{b1}M_{b2}$	$(1-r_a)(1-r_b)$	$r_a r_b$

$$= \prod_{i=1}^{n} \sum_{q=1} P_{mq} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i-\mu_q)^2}{2\sigma^2}\right] \tag{3}$$

其中 m_i 为个体或系 i 的标记基因型，共 n 个个体或系。

使该似然函数最大的 μ_j 和 σ^2 就是这些参数的最大似然估计值，为使参数 $\lambda \in (\mu_j, \sigma^2)$ 最大，先对似然函数取对数：

$$\ln L = \ln\left(\prod_{i=1}^{n} f(y_i \mid m_i;\ r_a)\right)$$

$$= \ln \frac{n}{\sqrt{2\pi\sigma^2}} + \sum_{i=1}^{n} \ln \sum_{q=1} P_{mq} \cdot \exp\left[-\frac{(y_i-\mu_q)^2}{2\sigma^2}\right] \tag{4}$$

如果定义

$$W_q(y_i \mid m_i; r_a) = \frac{P_{mq} f_q(y_i)}{f(y_i \mid m_i;\ r_a)} \tag{5}$$

则当某个体或系具有表型值 y，且标记基因型为 m 时，在该座位的QTL基因型为 q 的概率由 W_q 决定。令 (4) 式的微分为0，解之即得 λ 的估计值：

$$\hat{\mu}_q = \frac{\sum_{i=1}^{n} [W_q(y_i \mid m_i;\ r_a) \cdot y_i]}{\sum_{i=1}^{n} W_q(y_i \mid m_i;\ r_a)} \tag{6a}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\sum_{q=1}^{} [\, W_q(y_i \mid m_1; r_s) \cdot (y_i - \mu_q)^2 \,] \qquad (6b)$$

当 QTL位于两个标记座位之间时，(6)式不存在显式解。似然方程(6)的解采用EM迭代法(Dempster等 1977)获得。EM法的 E步骤就是根据已知数据和 $\lambda \in (\mu_1, \mu_2, \mu_3, \sigma^2)$ 的初始近似值（分别用标记基因型的数量性状平均值 x_1、x_2和 x_3以及群体样本方差 s^2代替），对未知的缺失数据求期望。M步骤是根据公式(6)，利用 λ 的初始值和E步骤得到的缺失数据的期望值，使对数似然函数(4)最大化，求出新一轮的 λ 值。E步骤和M步骤交替进行，每次用新的 λ 估值代替旧的 λ 估值，直到似然函数(4)不再增值(两次迭代的值小于预定的临界值 T) 时停止。

8.1.3 *LOD* 检验

当零假设 H_0: $\mu_i = \mu_L$ $(i \neq L)$ (即不存在QTL效应)成立时，似然函数变为:

$$L_0 = L(\mu_p, \sigma_p^2, y_1, y_2 \cdots y_n) = \prod_{i=1}^{n} f(y_i) \qquad (7)$$

其中:

$\hat{\mu}_p = \frac{1}{n}\sum_{i=1}^{n} y_i$ 为作图群体的平均数

$\hat{\sigma}_p^2 = \frac{1}{n}\sum_{i=1}^{n} (y_i - \mu_p)^2$ 为作图群体的方差

$$f(y_i) = \frac{1}{\sqrt{2\pi\,\sigma_p^2}} \exp\left[-\frac{(y_i - \mu_p)^2}{2\sigma_p^2}\right]$$

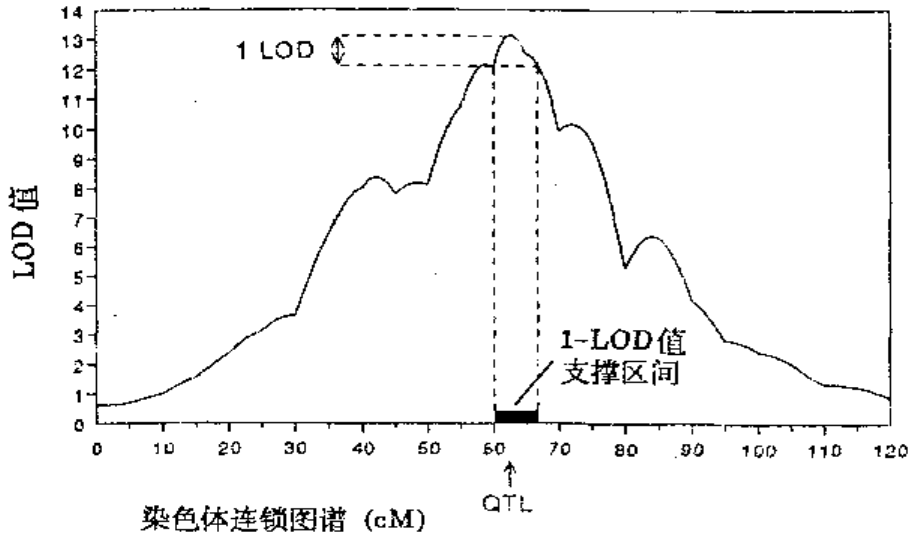


图 8.1　QTL似然图谱及其解释

即平均值为 μ_p，方差为 σ_p^2 的正态概率密度函数。备择假设（在该位置 QTL 处于分离状态）的似然比检验统计量可转换成 *LOD* 值

$$LOD=\log_{10}\left[\frac{L(\mu_j,\ \sigma^2, y_1,\ y_2,\cdots\ y_n)}{L_0(\mu_p, \sigma_p^2, y_1,\ y_2,\cdots\ y_n)}\right]$$

对两个标记座位 M_a 和 M_b 所包括的区间每隔 z cM 计算出相应的 *LOD* 值。将同一连锁群上所有标记间的 *LOD* 值连接起来即得到反映 QTL 可能位置的 QTL 似然图谱。

QTL 位置可估计为似然图谱上曲线最高处所对应的点。由比最大 *LOD* 值小 1 个单位的直线将在曲线上产生两个交点。交点所括区间即为 1 个 *LOD* 值支撑区间，即 QTL 存在位置的置信区间（图 8.1）。采用同样的方式，可以确定任意 *LOD* 值支撑区间。

8.2 区间作图软件

8.2.1 QTLMAP-Ⅰ

根据以上介绍的遗传学和统计学的原理，我们编写了适应于 QTL 区间作图的计算程序 QTLMAP-Ⅰ。该程序为 F_2、BC 和 DH 群体的通用程序，程序设计步骤如图 8.2。全部计算可在一般 PC286 及其以上的微机上完成。采用电子表格系统 Quattro Pro 3.0 对计算所得数据进行 QTL 似然图谱制作。

8.2.2 Mapmaker/QTL

我们以 MAPMAKER/QTL 1.1（Lincoln 等 1993）为例说明其操作过程。

1. 将数据调入 MAPMAKER/QTL

MAPMAKER/QTL 软件运行需要提供以下信息：(1) 确定所要检测 QTL 的染色体；(2) 将遗传标记定位在相应的染色体上；(3) 确定标记在染色体上的位置。可以利用 MAPMAKER/EXP 或其它的编辑软件来产生一个文件，供 MAPMAKER/QTL 调用。

2. 起动 MAPMAKER/QTL

Lincoln 等 (1993) 所编写的 MAPMAKER/QTL 1.1 具有可分别在 PC386 或 486 计算机 (DOS5.0)、苹果机 (A/UX 3.0, Apple's Unix)、Sun SPARCStation (SunOS 4.x) 上运行的软件。具体进入 MAPMAKER/QTL 的方式取决于所使用的计算机及软件是如何安装的。大多数 MAPMAKER/QTL 操作过程的第一步是调入数据文件。在软件所提供的实例中，数据文件分为两个，一是 sample.data，存放分子标记数据，一是 sample.traits，存放数量性状值。采用 load data 命令将这些文件调入 MAPMAKER/QTL。可采用 photo 命令来建立一个 qtl.out 文件以存放有关运算结果。

3. 数量性状数据的初步考察

MAPMAKER/QTL 的算法是建立在群体中数量性状值的变异服从正态分布这一假定上的。因此，大多数 QTL 作图的第一步是对所要分析的性状进行考察，以确定是否符合这一假定。对于不符合正态分布的性状，通常可以经过数据转换使之与假定更为符合。利用 trait 命令可选择所要分析的性状，然后利用 show trait 命令来显示该性状的一些统计量（均值，标准差，峰度和偏度等），据此可以判断该性状是否符合正态分布。对于非正态

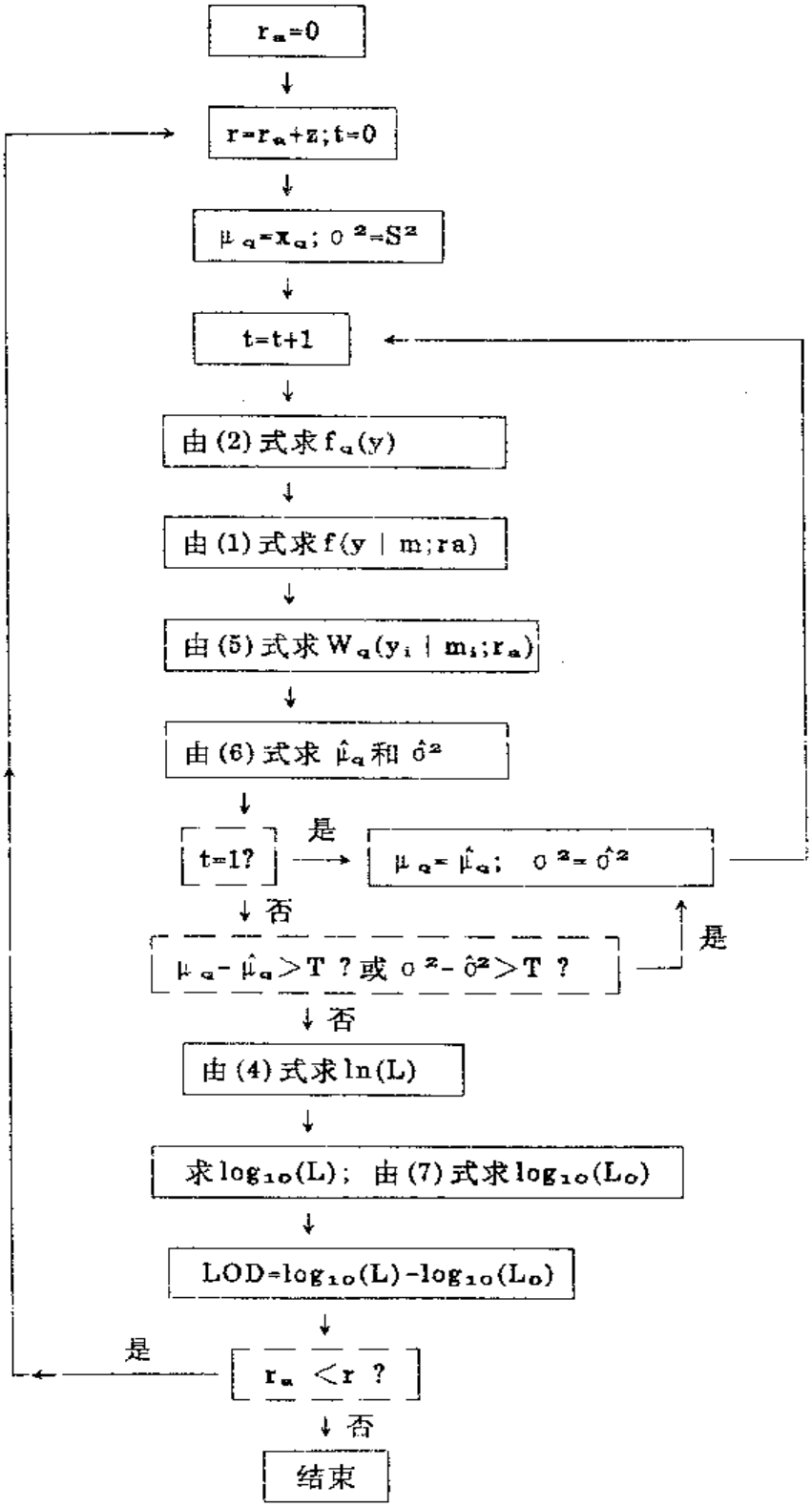


图 8.2　QTL 区间作图法求 LOD 值的计算机程序框图

分布的性状可通过适当的数学方法进行转换。例如，实例中的第一性状 (weight) 的分布偏向于高值方向，我们可以用分析 $\log_{10}$ (weight) 代替对 weight 的分析，这一转换与控制植株大小的 QTL 对重量的影响是指数增减的假定相对应。采用 make trait 命令，即可产生一个新的性状，使其值与原性状值之间存在某种函数关系。

4. 在基因组上检测 QTL

使用 sequence 命令确定所要检测的基因组区域，trait 命令确定所要分析的性状，scan 命令检测 QTL。在一般缺省命令下，在基因组上检测 QTL 的间距为2cM。scan 命令的输出结果包括 QTL 的效应、QTL 所能解释的变异占总变异的比例(等价于线性回归分析中的 r^2)、LOD 值等。利用 draw scan 命令还可以将 LOD 值用曲线图表示。根据所分析的基因组大小(范围)和性状的多少，scan 命令可能会要在计算机上运行相当长的时间。为了简便起见，也可采用 show peaks 命令显示出基因组上特定区域范围内最大 LOD 值所在的位置。

5. 给 QTL 拟合适当的遗传模型

采用 MAPMAKER/QTL，可以为所检测到的 QTL 拟合四种不同的遗传模型。对于 F_2 群体，任一 QTL 存在三种可能的基因型 A/A，A/B 和 B/B。个体 i 的数量性状表型值可用模型表述为：

$$Trait_i = Mean + (Weight \times Num_i) + (Dominace \times Het_i) + Noise$$

其中

Mean = 不被 QTL 效应所控制的性状均值部分，它等于所有 A/A 个体的性状均值；

Weigth = QTL B 等位基因效应的加性分量 (即 B等位基因的加性效应)；

Num_i = 个体 i 所含有的B等位基因数，其值为 0，1或2；

Dominance = QTL B 等位基因效应的显性分量 (B等位基因的显性效应)；

Het_i = 个体 i 为杂合体 A/B 时等于1，否则为零；

Noise = 不受该 QTL 所控制的变异 (为正态随机变量)

四种遗传模型为：

(1) Free 模型：A/A 效应 = 0

→ $Trait_i = Mean + (Weight \times Num_i) + (Dominace \times Het_i) + Noise$

(2) 加性模型：A/A 效应=0 且 B/B效应 = 2× A/B 效应

→ $Trait_i = Mean + (Weight \times Num_i) + (0 \times Het_i) + Noise$

→ dominance = 0

(3) 显性模型：A/A 效应 = 0 且 B/B 效应 = A/B 效应

→ $Trait_i = Mean + (Weight \times Num_i) + (Weight \times Het_i) + Noise$

→ dominance = weight

(4) 隐性模型：A/A 效应 = 0 且 A/B 效应 = A/A 效应

→ $Trait_i = Mean + (Weight \times Num_i) + (-Weight \times Het_i) + Noise$

→ dominance = - weight

改变 sequence 命令 [ ] 中内的内容，即可分别对上述四种模型进行拟合。详见

Lincoln 等 1993)。

6. 同步拟合多个 QTL

MAPMAKER/QTL 算法建立的一个基本假定是数量性状表现型由单一 QTL 所控制。如果存在两个连锁的 QTL，进行 scan 分析时就会看到两个相隔很近的 LCD 值峰。由于此时单个 QTL 的假定不能满足，因此一个 QTL 的分析就会受到与之连锁的另一QTL的影响，其结果将导致 QTL 位置的偏移及LOD 值的增加。而且，scan 分析得到的 LOD值的双峰分布并不一定就意味着存在两个 QTL。为此，MAPMAKER/QTL软件还提供了一种同步拟合多个 QTL 的方法，从而使其可用于分析表现可加性作用的两个独立的 QTL。两座位模型最为可能的参数包括：

$$\begin{aligned} Trait_i = Mean &+ (Weight_1 \times Num_{1i}) + (Dominace_1 \times Het_{1i}) \\ &+ (Weight_2 \times Num_{2i}) + (Dominace_2 \times Het_{2i}) + Noise \end{aligned}$$

8.3 应用

8.3.1 应用

根据本程序对郑康乐等(1992)表1所列数据，采用 QTLMAP-I 计算程序求得该连锁群上每隔 1 cM处的 LOD 值。借助于 Quattro Pro 3 电子表格系统绘制 QTL 似然图谱。结果表明，QTL 似然图谱在第6 染色体上色素原基因 *C* 和 DNA探针 RG138 之间有一个峰，最大 LOD 值(8.73) 出现在距 *C* 3 cM处 (图8.3)。这就是广亲和基因 S_a 最可能的染色体位置。

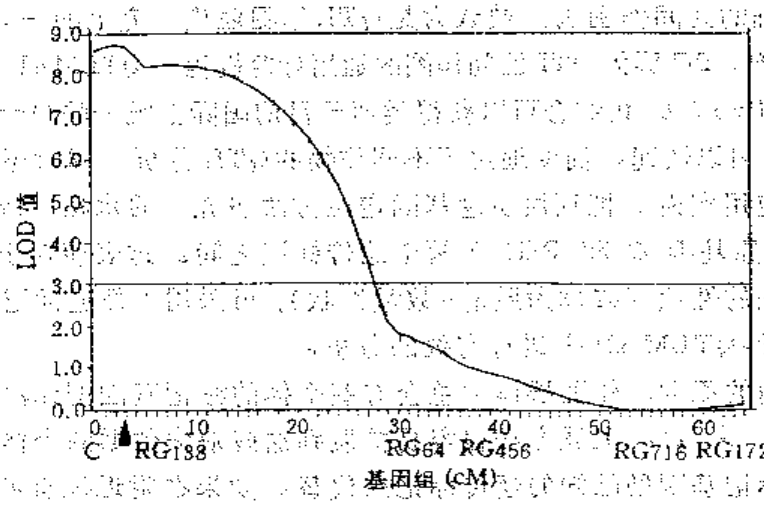


图 8.3 水稻第6染色体上结实率(广亲和性)的QTL 似然图谱

水稻籼粳杂交组合窄叶青 8 号/京系 17 F_2 群体 76 个单株的 DNA分别与第1染色体上的4个 RFLP 探针 (RG101、RG220、RG780 和RG462) 进行分子杂交和放射自显影(方法

参见朱立煌等1992)，得到各个体 RFLP 标记基因型。同时观察各单株的抽穗期(从播种至抽穗的天数）和每穗实粒数。 所得 QTL 似然图谱表明，每穗实粒数的最大 LOD 值(2.20)出现在距 RG101 4 cM处，此处可能存在一个与该性状有关的 QTL。抽穗期的QTL 似然图谱的峰值位于距 RG780 4 cM处（图 8.4)，最大 LOD 值仅1.52，表明在所研究的基因组范围内不存在与抽穗期有关的 QTL。

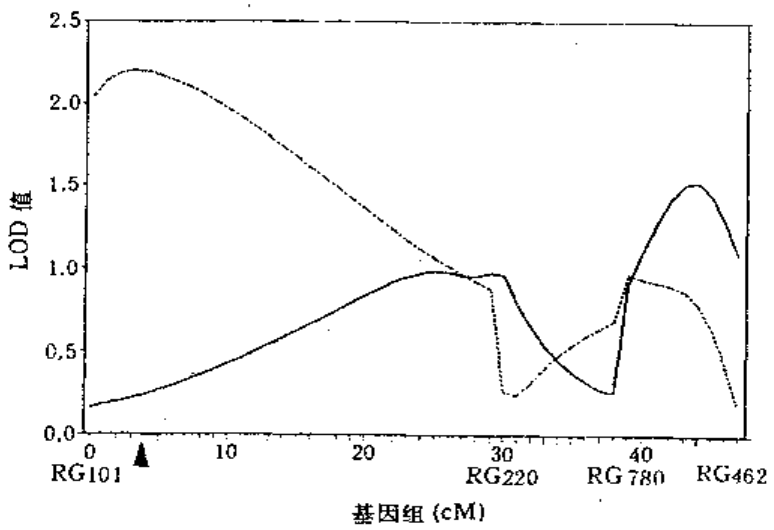


图 8.4　水稻第1染色体上抽穗期（实线）和每穗粒数（虚线）的QTL 似然图谱

上述两例应用中，3 个最大LOD 值有 2 个出现在两个侧连标记之间，距离标记1cM以上，表明区间作图法的确可以发现位于两个标记之间的 QTL。

以侧连标记为基础的区间作图法，被认为是QTL作图理论研究上的一大突破。我们自行编制了适合于F_2、DH、BC 群体 QTL区间作图的通用计算机程序 QTLMAP-I，该程序与电子表格系统 Quattro Pro 3.0 相结合可以获得类似于目前国际上唯一的Mapmaker /QTL 软件所能得到的结果。可以快速、简便地用于不同作物的QTL 分析。同时还为这类方法提供了两个不同群体的应用实例。把利用质量基因定位方法研究过的水稻广亲和基因 *Sg* 定位于第6染色体上色素原基因 *C* 和 RG138 两个遗传标记之间。这表明对于结实率之类兼有质量和数量遗传特征的性状（或称为质量-数量性状)，可采用以侧连标记为基础的区间作图法及QTL 作图软件QTLMAP-I 进行有效的分析。

由于实验技术方面的原因，作图群体中总会有些个体的标记基因型无法确定（参见表6.12及郑康乐等(1992) 表 1)。上述方法很容易扩展到这些缺失数据的个体。方法之一是采用附近紧密连锁且标记基因型已知的遗传标记来代替。如果在附近没有可供利用的侧连标记，则在 QTL 作图中可以不用这个个体，或采用与标记无关的 QTL 期望比例作为混合比例 P_{m1}、P_{m2} 和P_{m3}，即对于F_2 为0.25:0.50:0.25，BC为0:0.50:0.50，DH 为0.50:0:0.50。弃除标记基因型缺失的个体将使 LOD 值降低。同时，由于缺失数据因标记而不同，将对染色体不同部位的 LOD 值的可比性产生影响，除非缺失的数量不是太多。

LOD值反映了QTL存在或不存在于某一位置两种可能性的相对大小。由于LOD值大

小与作图群体的种类、样本大小及遗传图谱的饱和程度等有关，目前还没有适当的显著性测验标准。van Ooijen (1992)模拟研究了最大 LOD 值的概率分布，并考虑到所研究物种基因组的染色体数目。对于具有 12 条染色体的物种，要获得0.05的总显著水平，则每个独立的染色体的显著水平应为0.0043，F_2 和 BC群体相应的LOD临界值应分别为 3.7 和3.0。在本章所提供的实例中，有关广亲和基因的最大LOD 远超过了相应的临界值。

8.3.2 实例

我们以第5染色体上的第6、7两个标记(RG435和RZ225) 为例，说明采用区间作图法检测水稻ZYQ/JX F_2群体中与结实率有关的数量性状基因。这两个标记间的图距为15.3 cM。F_2群体结实率均值为49.96338，方差为373.8075，进而得到$\log_{10}(L_0)=-140.7895$。

对于两个标记座位 RG435和 RZ225所包括的区间每隔1cM计算出相应的 *LOD* 值。采用 EM 迭代法求似然方程(6)的解。

当距离标记RG435为 $\theta=1$cM时，经过双曲正切函数所转换成的重组率为 $r_m=9.998666$ E-03。由此计算出标记 RG435、RZ225 和假定的 QTL 三个基因座位所决定的27种基因型的频率如表8.4 (此表与表8.1相对应)。

表 8.4 两个侧连标记和 QTL 所确定的27种基因型的频率

标记座位		QTL 基因型频率		
RG435	RZ225	Q_1Q_1	Q_1Q_2	Q_2Q_2
1	1	2.749262E-06	3.310681E-03	0.9966865
1	2	9.997332E-05	1.979739E-02	0.9801027
1	3	3.348081E-03	0.109029	0.8876228
2	1	1.994374E-02	0.2425571	0.7374991
2	2	0.25	0.5	0.25
2	3	0.7374991	0.2425571	1.994374E-02
3	1	0.8876228	0.109029	3.348081E-03
3	2	0.9801027	1.979739E-02	9.997332E-05
3	3	0.9966865	3.310681E-03	2.749262E-06

根据表8.4，可得出作图群体中某个体 i 在 r_m 所确定的基因座位上 QTL 基因型为 q 的概率 $[W_q(y_i \mid m_i; r_m)]$。例如，第1号植株分子标记 RG435和 RZ225的基因型均为ZYQ 类型 (代码为1)，因而三种QTL基因型的频率分别为2.749262E-06,3.310681E-03和 0.9966865 。群体中 1～10号植株的 QTL 基因型频率如下：

植株号	标记基因型		QTL 基因型频率		
	RG435	RZ225	Q_1Q_1	Q_1Q_2	Q_2Q_2
1	1	1	2.749262E-06	3.310681E-03	0.9966865
2	3	3	0.9966865	3.310681E-03	2.749262E-06
3	1	2	9.997332E-05	1.979739E-02	0.9801027
4	1	1	2.749262E-06	3.310681E-03	0.9966865
5	3	3	0.9966865	3.310681E-03	2.749262E-06
6	2	2	0.25	0.5	0.25
7	2	2	0.25	0.5	0.25
8	3	3	0.9966865	3.310681E-03	2.749262E-06
9	2	3	0.7374991	0.2425571	1.994374E-02
10	3	3	0.9966865	3.310681E-03	2.749262E-06

将群体中各个体的 $W_q(y_i \mid m_i; r_a)$ 值，数量性状表型值代入 (6a) 即得 $\hat{\mu}_q$ (分别为 45.75626, 51.24739 和 54.11351)，将 $\hat{\mu}_q$ 代入 (6b) 即得 $\hat{\sigma}^2$ (值为 359.7582)。将这些估计值作为初始值代入(5)式，获得 $W_q(y_i \mid m_i; r_a)$ 的第一轮估计值。同上得到第一轮 $\hat{\mu}_q$ 和 $\hat{\sigma}^2$。将 $\hat{\mu}_q$ 和 $\hat{\sigma}^2$ 代入 (3) 式并取 $\log_{10}$，得到对数似然函数值 $\log_{10}(L)$，此值为 -139.9531。此过程进行下去，将使 $\log_{10}(L)$ 的值不断增加。当前后两次迭代的 $\log_{10}(L)$ 值之差小于0.0001时，停止迭代，由 $\log_{10}(L) - \log_{10}(L_0)$ 求出LOD 值。迭代结果如下：

轮 次	QTL 基因型均值			$\hat{\sigma}^2$	$\log_{10}(L)$
	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$		
1	44.62886	51.75643	55.08826	351.6539	-139.9531
2	44.29904	52.00542	55.31548	349.0802	-139.9478
3	44.19209	52.14785	55.35275	348.295	-139.9471
4	44.15222	52.23701	55.34526	348.0442	-139.9469
5	44.13448	52.29531	55.33091	347.954	-139.9468

当距离标记RG435为 θ =2cM时，经过双曲正切函数所转换成的重组率为 r_a =1.998934E-02。由此计算出标记 RG435、RZ225 和 QTL 三个基因座位所决定的27 种基因型的频率如表8.5 (此表与表8.1相对应)。

表 8.5　两个侧连标记和 QTL 所确定的27种基因型的频率

标记座位		QTL 基因型频率		
RG435	RZ225	Q_1Q_1	Q_1Q_2	Q_2Q_2
1	1	9.856933E-06	6.259437E-03	0.9937307
1	2	3.995737E-04	3.917953E-02	0.9604208
1	3	1.361535E-02	0.206139	0.7802457
2	1	1.789037E-02	0.231729	0.7503806
2	2	0.25	0.5	0.25
2	3	0.7503806	0.231729	1.789037E-02
3	1	0.7802457	0.206139	1.361535E-02
3	2	0.9604208	3.917953E-02	3.995737E-04
3	3	0.9937307	6.259437E-03	9.856933E-06

同理，根据表8.5，可得出作图群体中各个体在 r_m = 1.998934E-02时所确定的基因座位上 QTL 基因型为 q 的概率 $[W_q(y_i \mid m_i; r_m)]$。进而将群体中各个体的 $W_q(y_i \mid m_i; r_m)$ 值，数量性状表型值代入(6a)即得 $\hat{\mu}_q$（分别为45.63712，51.61858和53.98911），将 $\hat{\mu}_q$ 代入 (6b) 即得 $\hat{\sigma}^2$（值为 359.5534）。将这些估计值作为初始值代入(5)式，获得 $W_q(y_i \mid m_i; r_m)$ 的第一轮估计值。同上得到第一轮 $\hat{\mu}_q$ 和 $\hat{\sigma}^2$。将 $\hat{\mu}_q$ 和 $\hat{\sigma}^2$ 代入 (3) 式并取$\log_{10}$，得到对数似然函数值 $\log_{10}(L)$，此值为 -139.9159。各轮迭代结果如下：

轮 次	QTL 基因型均值			$\hat{\sigma}^2$	$\log_{10}(L)$
	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$		
1	44.38599	52.40084	54.90551	350.7575	-139.9195
2	43.97953	52.85134	55.08083	347.6507	-139.9104
3	43.82733	53.13913	55.07368	346.5213	-139.9082
4	43.75983	53.33146	55.03118	346.0459	-139.9073
5	43.72436	53.46232	54.99078	345.8047	-139.907
6	43.70321	53.55204	54.95969	345.6621	-139.9068
7	43.6896	53.61366	54.9373	345.5696	-139.9067

标记 RG435和 RZ225 之间其它位置（距离标记 RG435 为1～15cM）处，所得 LOD 值列于表 8.6 。由此表可见，在距标记 RG435 10 cM处，LOD 达到最大(1.114883)。

表 8.6　利用标记 RG435 和RZ225 对结实率 QTL 进行区间作图的结果
[$\log_{10}(L_0) = -140.7895$]

染色体位置（距RG435的cM数）	迭代次数	$\log_{10}(L)$	LOD
1	5	-139.9468	0.8427
2	7	-139.9067	0.8828
3	8	-139.8614	0.928131
4	9	-139.8163	0.973190
5	10	-139.7749	1.014633
6	10	-139.7390	1.050568
7	10	-139.7098	1.079712
8	10	-139.6886	1.100876
9	10	-139.6765	1.113022
10	11	-139.6746	1.114883
11	11	-139.6844	1.105072
12	11	-139.7075	1.082047
13	10	-139.7451	1.044464
14	9	-139.7979	0.991608
15	8	-139.8649	0.924576

对于所研究的5个数量性状，第5染色体上各成对标记间的 最大 LOD 值及与左侧标记的距离列于表 8.7。其中，在标记 RZ70和 RG697 之间与颖花数有关的LOD 值最高，达 2.79，表明存在 QTL 的可能性较大。

表8.7　ZYQ/JX F2 群体每一标记区间的最大LOD值及其位置(以离左侧标记的距离cM表示

标记区间及图距	n	特征参数	生育期	株高	穗长	颖花数	结实率
RG573——RG13 16.5cM	75	位置(cM)	16	1	1	1	1
		最大LOD值	0.68	0.81	1.45	1.54	0.16
RG13——RZ70 14.7cM	75	位置(cM)	2	14	14	14	14
		最大LOD值	0.47	1.64	2.35	1.83	0.30
RZ70——RG697 26.8cM	67	位置(cM)	13	9	1	7	23
		最大LOD值	0.78	1.49	2.12	2.79	1.65
RG697——RG474 17.4cM	66	位置(cM)	17	17	17	17	1
		最大LOD值	0.43	1.07	0.30	0.47	1.71
RG474——RG435 10.1cM	74	位置(cM)	1	1	1	1	10
		最大LOD值	0.29	0.49	0.06	0.27	1.13
RG435——RZ225 15.3cM	74	位置(cM)	1	15	15	15	10
		最大LOD值	0.24	1.28	0.19	0.29	1.11

第 9 章 QTL作图：复合区间作图法

现已发展了多种统计方法来分析实验生物的作图数据，以系统地检测主要的QTL（如Soller等1976；Weller 1986；Lander和Botstein 1989），其中Lander和Botstein的方法现已成为许多遗传学家用于 QTL作图的标准方法。与传统方法（Soller等1976）相比，区间作图法具有一些优点，但仍然存在一些问题。这些问题包括：(1) 某一区间的检验统计量可能受到位于该染色体上其它区域的 QTL的影响。既使在某一区间不存在QTL，但如果在该染色体上其附近存在一个 QTL，则该区间的检验统计量仍有可能极为显著。而且，如果在某一染色体上存在一个以上的 QTL，则正在检验的位置的统计量就会受到所有这些 QTL的影响，且由此所估计的位置及所鉴别的 QTL 效应可能是有偏的（Haley和Knott 1992；Martinez和Curnow 1992）。(2) 考虑到来自其它标记的信息，一次只应用两个标记进行检验不是一种有效的方法。Lander和Botstein (1989) 曾倡导了对多个区间上的多个QTL进行同步检测的策略以减轻这些问题。但由于这种检测涉及到多维空间（Lander和Botstein 1989；Haley和Knott 1992），因而在参数估计及模型鉴别上存在一些困难。此外，一个染色体上QTL 的数目是未知的，作图仍然是有偏的，而且仍然没有利用其它标记所提供的信息。

理想的情形应当是：在检验某一区间的 QTL时，其检验统计量与该染色体上其它区域可能存在的 QTL的效应是独立的。这样，我们就可以消除来自这些 QTL的偏倚作用，从而提高作图的精度。这样一种统计量可以将 Lander和Botstein 的区间作图与多元回归分析相结合来构建。Zeng(1993) 探讨了将多元回归分析用于 QTL作图的理论问题。结果表明，多元回归中表型对标记的偏回归系数只取决于两个相邻标记所包括的区间所存在的QTL，且与其它区间的 QTL无关。因此，应用这一特性，我们可以构建与该染色体上其它QTL的效应无关的检验统计量。这一结果为 QTL 作图的区间检验提供了基础。而且，通过在回归模型中拟合多个标记，在分析中就可控制群体中的大量背景遗传变异，从而提高检测 QTL的统计效率。Zeng (1993) 还讨论了在QTL作图中给该模型拟合多个标记时的优点和缺点。

为提高 QTL作图的可靠性和准确性，在检验和估计中应当将同一染色体上的多个连锁QTL 的效应区分开来。Zeng(1994) 发展一种作图方法，利用多个标记以改善多个QTL 作图的精度和效率。这一方法的基础是区间检验，所构建的检验统计量不受区间以外的其它QTL 的影响。这是通过利用多元回归分析的性质（Zeng 1993）来实现的。尽管多元回归分析已用于QTL作图（Cowen 1989），但只是在最近，Zeng(1993)，Rodolphe和Lefort (1993) 才详细地分析了与 QTL作图有关的多元回归分析的理论性质。Zeng (1993) 还建议利用这些性质发展复合区间作图方法以提高QTL 作图的精度和效率。最近，Jansen (1993) 也发展了一种将多元回归分析与区间作图相结合的 QTL作图方法。

本章首先提供有关的背景和问题并概括多元回归分析的性质，然后详细描述回交群体中的复合区间作图方法，讨论与检验的适当临界值确定有关的统计问题，最后提供QTL作图的模拟结果以说明这一方法的各种性质。

9.1 区分多个连锁QTL的理论基础

9.1.1 统计模型

为简便起见，我们考虑在 m 个QTL 和 t 个标记座位上不同等位基因固定的两个近交亲本 P_1和 P_2所产生的回交群体。假定P_1和 P_2 的群体平均数为 μ_1和 μ_2，μ_1和 μ_2 之间的差异为

$$\mu_1 - \mu_2 = \sum_{u=1}^{m} c_u \quad \text{（不考虑上位性效应）}$$

其中 c_u 为第 u 个 QTL处两个亲本群体中等位基因固定的两种纯合体之间的效应差值。c_u 可正可负。 两个亲本杂交所产生的F_1的均值定义为

$$\mu_{F1} = \mu_2 + \sum_{u=1}^{m} 1/2\ (1+\ d_u)\, c_u$$

其中 d_u 为第 u 个QTL的显性度。 P_1和 F_1 之间回交群体中各个体的性状值定义为

$$y_h = \mu_2 + \sum_{u=1}^{m} [\ \xi_{uh} c_u + 1/2(1-\xi_{uh})(1+d_u)\, c_u\ ] + e_h$$

其中y_h 为第h 个体的性状值，ξ_{uh} 为具有相等概率取值为1和0的指标变量，e_h 为具有平均值为0、方差为 σ_e^2的随机环境离差。如果第 u 和第 v 个QTL是不连锁的，则 ξ_{uh} 和 ξ_{vh} 相互独立，否则， ξ_{uh} 和 ξ_{vh} 是相互相关的，相关系数为

$(1-2\ r_{uv})$

其中 r_{uv} 为第 u 和 v 两个QTL之间的重组率。则回交群体中性状值的方差为:

$$\sigma_y^2 = \sigma_e^2 + 1/4 \sum_{u=1}^{m} a_u^2 + 1/4 \sum_{v=1}^{m} \sum_{u=1,\ u\neq m}^{m} (1-2r_{uv})\, a_u a_v$$

其中 $a_u = (1-\ d_u)\ c_u/2$为用P_1纯合体和F_1杂合体之间的效应之差表示的第 u 个QTL的效应。和的两倍为基因座位间由于连锁不平衡所引起的协方差。

假定回交群体的样本数为 n，并对数量性状进行观察，且获得了 t 个排序的标记。分析这些数据的一种方式就是利用下列线性模型进行多元回归分析:

$$y_h = b_0 + \sum_{i=1}^{t} b_i x_{hi} + \zeta_h \quad (h=1,\ 2,\ \ldots,\ n)$$

其中 x_{hi} 为第 h 个体第 i 标记的类型，当标记为纯合时取值为1，为杂合时取值为0；b_0为该模型的均值；b_i （也用 $b_{yi\cdot s_i}$ 表示，s_i 代表除第i标记外所有其它标记组成的集合）为表型y在所有其它标记存在的条件下对第i标记的偏回归系数；ζ_h 为随机变量。在这里，我们用下标 u、v、q 用于QTL，i、j、k 和 l 用于标记，h 用于个体。

但是这一分析本身并不适合于 QTL作图。不过，多元回归分析有几个独特的性质，

可用于发展 QTL作图方法，从而提高 QTL作图的精度和效率。下面我们将分析回归系数的期望值、系数的期望抽样方差、系数的抽样相关，以探讨多元回归分析的优点和缺点。

9.1.2 偏回归系数

为了分析偏回归系数 $b_{yi.wi}$ $(= \sigma_{yi.wi} / \sigma^2_{i.wi})$，首先分析条件方差 $\sigma^2_{i.wi}$ 和条件协方差 $\sigma_{yi.wi}$。由于是回交群体，群体中每一标记变量的方差为

$$\sigma_i^2 = 1/4 \quad (i = 1, 2, \ldots, t)$$

很容易证明第 i 和 j 个标记间的协方差为

$$\sigma_{ij} = (1-2r_{ij})/4$$

其中 r_{ij} 为这两个标记之间的重组率。 性状值 y 与第 i 标记之间的协方差为

$$\sigma_{yi} = \sum_{j=1}^{m} (1-2r_{uj}) \ a_u \ /4$$

其中 r_{ui} 为第 u QTL与第 i 标记间的重组率。

利用这些基本公式，可推导出任何的条件方差和协方差(Stuart和Ord1991)。首先，在 Haldane作图函数(假定无干扰)时，标记 i 和 j 之间的重组率可表示为

$$(1-2r_{ij}) = (1-2r_{ik})(1-2r_{kj}) \quad (i<k<j)$$

为简便起见，假定基因组上的所有标记都是按从第一染色体到最后一个染色体的物理顺序编号的，因此顺序 $i<k<j$ 是指标记 k 位于标记 i 和 j 之间。对于不同染色体上的两个标记，这种相对顺序是不重要的。

在标记 j 条件下标记 i 的方差为

$$\sigma^2_{i.j} = \sigma_i^2 - \sigma_{ij}^2 / \sigma_j^2 = [1 - (1-2r_{ij})^2]/4 = r_{ij}(1-r_{ij})$$

在标记 k 条件下的标记 i 和 j 之间的协方差为

$$\sigma_{ij.k} = \sigma_{ij} - \sigma_{ik}\sigma_{jk} / \sigma_k^2 = [(1-2r_{ij}) - (1-2r_{ik})(1-2r_{jk})]/4$$

当 $i<k<j$ 或 $j<k<i$ 时，上式为0;

当 $i<j<k$ 或 $k<j<i$ 时，上式为 $r_{jk}(1-r_{jk})(1-r_{ij})$;

当 $k<i<j$ 或 $j<i<k$ 时，上式为 $r_{ik}(1-r_{ik})(1-r_{ij})$。

对于 $i<j$ 和 $k<l$，在标记 k 和 l 条件下标记 i 和 j 之间的协方差为

$$\sigma_{ij.kl} = \sigma_{ij.k} - \sigma_{il.k}\sigma_{jl.k} / \sigma^2_{l.k}$$

$$= \sigma_{ij.l} - \sigma_{ik.l}\sigma_{jk.l} / \sigma^2_{k.l}$$

当 $i<k<j<l$ 或 $i<k<l<j$ 或 $k<i<j<l$ 时，上式为 0;

当 $i<j<k<l$ 时，上式为 $\sigma_{ij.k}$;

当 $k<l<i<j$ 时，上式为 $\sigma_{ij.l}$;

当 $k<i<j<l$ 时，上式为 $[r_{ki}(1-r_{ki})r_{jl}(1-r_{jl})(1-r_{ij})]/[r_{kl}(1-r_{kl})]$

从以上分析可知，如果 $k<i<j<l$，在标记 $k-1$、k、l、$l+1$ 标记下标记 i 和 j 之间的协方差为

$$\sigma_{ij.(k-1)kl(l+1)} = \sigma_{ij.kl(l+1)} = \sigma_{ij.kl}$$

因此一般而言，在所有其它标记条件下标记 i 和 $i+1$ 之间的协方差为

$$\sigma_{i(i+1).wi(i+1)} = \sigma_{i(i+1).(i-1)(i+2)k}$$

其中 $S_{i(i+1)}$ 为除标记 i 和 $i+1$ 外的所有标记的集合。

非常重要的是，在以中间标记为条件时，两个侧连标记之间的协方差为0。例如，如果 $i<k<j$，则 $\sigma_{ij.k}=0$。另外当以某一侧连标记（比如 k）为条件时，标记 i 和 j 之间的协方差就会独立于所有位于标记 k 的另一端的其它标记（和QTL，见下文），例如当 $i<j<k<l$ 时，$\sigma_{ij.kl}=\sigma_{ij.k}$。因此，当以两个侧连标记（如标记 $i-1$ 和 $i+2$）为条件时，两个内部标记 i 和 $i+1$ 之间的协方差就与标记区间（$i-1$，$i+2$）以外的所有标记（和QTL，见下文）无关。以上就是染色体上标记和QTL线性排列的特性，也正是区间检验的基础。 此外，考虑 $j<k$，在标记 j 和 k 条件下标记 i 的方差为

$$\sigma^2_{i.jk}=\sigma^2_{i.j}-\sigma^2_{ik.j}/\sigma^2_{k.j}$$
$$=\sigma^2_{i.k}-\sigma^2_{ij.k}/\sigma^2_{j.k}$$

其中 当 $i<j<k$ 时，上式为 $\sigma^2_{i.j}$

当 $j<k<i$ 时，上式为 $\sigma^2_{i.k}$

当 $j<i<k$ 时，上式为 $[r_{ji}(1-r_{ji})r_{ik}(1-r_{ik})]/[r_{jk}(1-r_{jk})]$

不难看出，一般而言，存在

$$\sigma^2_{i.ai}=\sigma^2_{i.(i-1)(i+1)}$$

在标记 j 条件下 y 和标记 i 之间的协方差较为复杂：

$$\sigma_{yi.j}=\sigma_{yi}-\sigma_{yj}\sigma_{xij}/\sigma^2_j$$
$$=1/4\sum_{u=1}^{m-1}[(1-2r_{ui})-(1-2r_{uj})(1-2r_{ij})]a_u$$
$$=\begin{cases}r_{ij}(1-r_{ij})\sum_{u\le i}(1-2r_{ui})a_u+\sum_{i<u<j}r_{uj}(1-r_{uj})(1-2r_{iu})a_u & (i<j)\\ r_{ij}(1-r_{ij})\sum_{u\ge i}(1-2r_{ui})a_u+\sum_{i<u<j}r_{ju}(1-r_{ju})(1-2r_{ui})a_u & (j<i)\end{cases}$$

显然，

$$\sigma_{yi.jk}=\sigma_{yi.j}-\sigma_{yk.j}\sigma_{ik.j}/\sigma^2_{k.j}$$
$$=\sigma_{yi.k}-\sigma_{yj.k}\sigma_{ij.k}/\sigma^2_{j.k}$$

其中 当 $i<j<k$ 时，上式为 $\sigma_{yi.j}$

当 $i<k<j$ 时，上式为 $\sigma_{yi.k}$

如果 $k<i<j$，则经过分析之后可以发现：

$$\sigma_{yi.jk}=\frac{r_{ij}(1-r_{ij})}{r_{kj}(1-r_{kj})}\sum_{k<u\le i}r_{ku}(1-r_{ku})(1-2r_{ui})a_u$$
$$+\frac{r_{ki}(1-r_{ki})}{r_{kj}(1-r_{kj})}\sum_{i<u<j}r_{uj}(1-r_{uj})(1-2r_{iu})a_u$$

一般而言，存在

$$\sigma_{yi.ai}=\sigma_{yi.(i-1)(i+1)}$$

因此偏回归系数为

$$b_{yi.ai}=\frac{\sigma_{yi.ai}}{\sigma^2_{i.ai}}=\frac{\sigma_{yi.(i-1)(i+1)}}{\sigma^2_{i.(i-1)(i+1)}}$$
$$=\sigma^2_{i.k}-\sigma^2_{ij.k}/\sigma^2_{j.k}$$

$$= \sum_{i-1<u\leqslant i} \frac{r_{(i-1)u}(1-r_{(i-1)u})(1-2r_{ui})}{r_{(i-1)i}(1-r_{(i-1)i})} a_u + \sum_{i<u<i+1} \frac{r_{u(i+1)}(1-r_{u(i+1)})(1-2r_{iu})}{r_{i(i+1)}(1-r_{i(i+1)})} a_u \quad (1)$$

其中，第一项求和是针对标记 $i-1$ 和 i 之间的所有 QTL，第二项求和是针对标记 i 和 $i+1$ 之间的所有 QTL。该回归系数只取决于标记 $i-1$ 和 $i+1$ 之间的 QTL。这是一个非常理想的特性。利用这一特性，可以构建一种区间检验的方法，对某一标记区间内是否存在QTL进行检验。与只考虑在某一染色体上是否存在 QTL 的非条件检验相比，这种条件检验更为精确。

如果在标记 $i-1$ 和 $i+1$之间只存在一个QTL（比如存在于标记 $i-1$和 i 之间），则

$$b_{yi.ei} = \frac{r_{(i-1)u}(1-r_{(i-1)u})(1-2r_{ui})}{r_{(i-1)i}(1-r_{(i-1)i})} a_u$$

当 $r_{(i-1)i}$ 较小（如（0.2）时，

$$b_{yi.ei} \approx [r_{(i-1)u} / r_{(i-1)i})] \quad a_u$$

除非QTL u 位于标记 i 右侧，$b_{yi.ei}$ 的估计值将是 a_u 的一个有偏估计值。随着QTL u 将其位置从标记 $i-1$ 移动到标记 i，比率 $r_{(i-1)u} / r_{(i-1)i}$ 将单调且几乎呈线性地从0变化为1。在统计上，这种条件检验就是使标记区间 $(i-1, i+1)$ 象一个染色体一样起作用。

9.1.3 偏回归系数的抽样方差

为分析偏回归系数的抽样方差，需要获得条件表型方差。Stuart和Ord (1991)给出的该方差为

$$\sigma_{y.e}^2 = \sigma_y^2 - \sum_{i=1}^{t} b_{yi.ei}\sigma_{yi}$$

$$= \sigma_y^2 - \sum_{i=1}^{t} [\{ \sum_{i-1<u\leqslant i} \frac{r_{(i-1)v}(1-r_{(i-1)v})(1-2r_{vi})}{r_{(i-1)i}(1-r_{(i-1)i})} a_v + \sum_{i<v<i+1} \frac{r_{v(i+1)}(1-r_{v(i+1)})(1-2r_{iv})}{r_{i(i+1)}(1-r_{i(i+1)})} a_v \} \frac{1}{4} \sum (1-2r_{ui}) a_u]$$

$$= \sigma_e^2 + \frac{1}{4}\sum_{u=1}^{m} a_u^2 + \frac{1}{4}\sum_{v=1}^{m}\sum_{u=1,u\neq v}^{m} (1-2r_{uv}) a_u a_v$$

$$- \frac{1}{4}\sum_{i=1}^{t} [\sum_{i-1<u\leqslant i} \frac{r_{(i-1)v}(1-r_{(i-1)v})(1-2r_{vi})^2}{r_{(i-1)i}(1-r_{(i-1)i})} a_v^2 + \sum_{i<v<i+1}^{t} \frac{r_{v(i+1)}(1-r_{v(i+1)})(1-2r_{iv})^2}{r_{i(i+1)}(1-r_{i(i+1)})} a_v^2$$

$$+ \sum_{i-1<v\leqslant i} \sum_{u=1,u\neq v}^{m} \frac{r_{(i-1)v}(1-r_{(i-1)v})(1-2r_{vi})(1-2r_{ui})}{r_{(i-1)i}(1-r_{(i-1)i})} a_u a_v$$

$$+ \sum_{i<v\leqslant i+1} \sum_{u=1,u\neq v}^{m} \frac{r_{v(i+1)}(1-r_{v(i+1)})(1-2r_{iv})(1-2r_{ui})}{r_{i(i+1)}(1-r_{i(i+1)})} a_u a_v]$$

$$= \sigma_e^2 + \frac{1}{4}\sum_{u=1}^{m}\left[1-\frac{r_{uLu}(1-r_{uLu})(1-2r_{uuR})^2 + r_{uuR}(1-r_{uuR})(1-2r_{uLu})^2}{r_{uLuR}(1-r_{uLuR})}\right]a_u^2$$

$$+\frac{1}{4}\sum_{u=1}^{m-1}\sum_{v=u+1}^{m}\left[2(1-2r_{uv}) - \frac{r_{uLu}(1-r_{uLu})(1-2r_{uuR})(1-2r_{uRv})}{r_{uLuR}(1-r_{uLuR})}\right.$$

$$-\frac{r_{uuR}(1-r_{uuR})(1-2r_{uLu})(1-2r_{uLv})}{r_{uLuR}(1-r_{uLuR})}$$

$$-\frac{r_{vLv}(1-r_{vLv})(1-2r_{vvR})(1-2r_{uvR})}{r_{vLvR}(1-r_{vLvR})}$$

$$\left.-\frac{r_{vvR}(1-r_{vvR})(1-2r_{vLv})(1-2r_{uvL})}{r_{vLvR}(1-r_{vLvR})}\right]a_u a_v \qquad (2)$$

其中 s 为所有标记的集合。在上式中，标记和QTL之间的连锁关系由 $\Sigma_{i<v<i+1}$ 之类的求和以及 u_L 和 u_R 之类的记号所表示，其中 $\Sigma_{i<v<i+1}$ 是指对标记 i 和 $i+1$ 之间的所有QTL求和，u_L 和 u_R 是指第 u 个 QTL的左侧和右侧的标记。公式(2)中的第一求和项包含了拟合所有可供利用的标记之后 QTL内部的剩余遗传方差；而第二求和项则包含了QTL之间的剩余遗传协方差。 如果每一标记区间最多只存在一个QTL（即任意两个QTL之间至少存在一个标记以至对于所有的 $u<v$，QTL u 的右侧标记总是在QTL v 的左边，也就是，$u_R<v$ 和 $u<v_L$），则第二求和项为0且在条件方差中将不存在QTL之间的剩余遗传协方差，因为这时所有的QTL都被标记所分开并使之相互独立。 因此

$$\sigma_{y.s}^2 = \sigma_e^2 + \sum_{u=1}^{m}\left[\frac{r_{uLu}(1-r_{uLu})r_{uuR}(1-r_{uuR})}{r_{uLuR}(1-r_{uLuR})}\right]a_u^2$$

如果标记在整个基因组上均匀分布，每一标记区间的重组率为 r（不包括连接两个染色体的隐含区间），则在所有QTL 都位于标记区间的中间（最糟的情形）的假定下，条件方差至多为

$$\sigma_{y.s}^2 = \sigma_e^2 + \sum_{u=1}^{m}\left[\frac{r}{4(1-r)}\right]a_u^2$$

与方差 σ_y^2 的遗传部分相比，方差 $\sigma_{y.s}^2$ 的遗传部分将至少减少因子数 $(1-2r)/r$。考虑到方差 σ_y^2 中的连锁不平衡，其减少的程度还将增加。

因此， 根据 (Stuart和Ord 1991)， 对于大样本， 回归系数的抽样方差近似为

$$\mathrm{Var}(\hat{b}_{yi.si}) \approx \frac{\sigma_{y.s}^2}{n\sigma_{i.si}^2}$$

$$= \frac{1}{n}\left[\sigma_e^2 + \sum_{u=1}^{m}\frac{r_{uLu}(1-r_{uLu})r_{uuR})(1-r_{uuR})}{r_{uLuR}(1-r_{uLuR})}a_u^2\right]$$

$$\Big/\ \frac{r_{(i-1)i}(1-r_{(i-1)i})r_{i(i+1)}(1-r_{i(i+1)})}{r_{(i-1)(i+1)}(1-r_{(i-1)(i+1)})} \qquad (3)$$

令人感兴趣的是比较公式(3)和线性模型 $y_h = b_0 + b_i x_{hi} + \zeta_h$ $(h=1, 2, \ldots, n)$ 中的简单回归系数的抽样方差。假定只有一个QTL q，与标记 i 相连锁，性状值 y 对标记 i 的简单回归系数的抽样方差 (Stuart和Ord 1991) 为

$$\mathrm{Var}(\hat{b}_{yi}) \approx \frac{\sigma_{y.i}^2}{n\ \sigma_i^2}$$

$$= \frac{4}{n}\Big[\ \sigma_e^2 + r_{iq}\ (1-r_{iq})\,a_q^2 + \frac{1}{4}\sum_{u=1,u\neq q}^{m} a_u^2 + \frac{1}{4}\sum_{u=1}^{m}\sum_{v=1,v\neq u}^{m}(1-2r_{uv})\,a_u a_v\ \Big]$$

由于 $\sigma_{y.a}^2 < \sigma_{y.i}^2$ 和 $\sigma_{i.ai}^2 \leqslant \sigma_i^2$，故 $\mathrm{Var}(\hat{b}_{yi.ai})$ 可能小于或大于 $\mathrm{Var}(\hat{b}_{yi})$，这取决于 $\sigma_{y.i}^2$ 比 $\sigma_{y.a}^2$ 所减少的程度及 $\sigma_{i.ai}^2$ 比 σ_i^2 减少的程度。对于多个QTL，

$\sigma_{y.a}^2$ 的遗传部分非常近似于 $\sigma_{y.i}^2$ 的遗传部分的 r 倍，其中 r 为标记区间的平均重组率。因遗传率的大小，这种减少可能非常明显。此外，如果 $r_{(i-1)i}<0.5$ 或 $r_{i(i+1)}<0.5$，$\sigma_{i.ai}^2$ 将小于 σ_i^2，结果 $\mathrm{Var}(\hat{b}_{yi.ai})$ 可能大于 $\mathrm{Var}(\hat{b}_{yi})$。

但是，由于 $\sigma_{i.ai}^2$ 只是相邻区间[如公式(3)中的区间$(i-1,\ i)$和$(i,i+1)$]的大小的函数，因此在其它标记的条件下有意去除一些连锁的标记(如去除标记 $i-1$ 和 $i+1$，这样 $\sigma_{i.a(i-1)(i+1)}^2$ 就是 $r_{(i-2)i}, r_{i(i+2)}$ 和 $r_{(i-2)(i+2)}$ 的函数)，将会使抽样方差(公式 3)的分母增加，同时最大限度地使该方差的分子减小，因而最大限度地减小偏回归系数的抽样方差，提高检验的统计效率。这是因为对 $\hat{b}_{yi.ai}$ 的条件检验取决于在标记 $i-1$ 和 i 之间或标记 i 和 $i+1$ 之间发生重组的重组体的数量。如果区间 $(i-1,i)$ 和 $(i,i+1)$ 较小，这类重组体就较少，用于检验的有效样本就会较小。因此，在检验中增加相邻标记区间的数目(以检验 $\hat{b}_{yia(i-1)i(i+1)}$ 为例)，将会增加重组体的数目，从而提高检验的统计效率。但是如果从模型中去除一些连锁的标记，如标记 $i-1$ 和 $i+1$，就会增加区间 $(i-2,i+2)$ 上可能存在的多个连锁的 QTL干扰假设测验和参数估计的机会。

9.1.4 偏回归系数的抽样相关

可以证明偏回归系数 $\hat{b}_{yi.ai}$ 和 $\hat{b}_{yj.aj}$ $(i<j$ 且假定 $r_{ij}>0)$ 的抽样相关为

$$\mathrm{Corr}\,(\hat{b}_{yi.ai}, \hat{b}_{yj.aj}) = -\gamma_{ij.aij} = -\frac{\sigma_{ij.aij}}{\sigma_{i.ai}\,\sigma_{j.aij}}$$

当 $j=i+1$ 时，上式等于

$$= -(1-2r_{i(i+1)})\sqrt{\frac{r_{(i-1)i}(1-r_{(i-1)i})\,r_{(i+1)(i+2)}(1-r_{(i+1)(i+2)})}{r_{i(i+2)}(1-r_{i(i+2)})\,r_{(i-1)(i+1)}(1-r_{(i-1)(i+1)})}} \qquad (4)$$

而在其它情形下，上式等于零。其中 $\gamma_{ij.aij}$ 是在所有其它标记条件下标记 i 和 j 之间的偏相关系数。这表明，表型对标记 i 和 j 的偏回归系数之间的抽样相关一般为0，除非这两个标记是相邻的。

这种相关与各个 $\hat{b}_{yi.ai}$ 和 $\hat{b}_{yj.aj}$ 的 F 检验统计量的相关存在密切关系。例如，如果要检验假设 H_0: $b_{yi.ai}=0$ 和 $b_{yi.ai}\neq 0$，就可以利用 F 统计量：

$$F_i = \hat{b}_{yi.ai}^2 \ /\ \mathrm{Var}(\hat{b}_{yi.ai})$$

其中 $\mathrm{Var}(\hat{b}_{yi.ai})$ 是 $\hat{b}_{yi.ai}$ 的抽样方差的一个估计值，统计量 F_i 具有自由度为1和 $n-t-1$ 的 F 分布。当样本容量 n 较大（显著大于拟合在模型中的标记数目 t）时，F 分布近似地服从自由度为1的 χ^2 分布。如果回归模型中的 ζ_h 服从正态分布，则 $\hat{b}_{yi.ai}$ 也服从(多元)正态分布，且具有公式 1、3、4 所给定的平均数、方差和相关。因此，根据正态分布的性质，当 $n \to \infty$ 时，F_i 和 F_j 之间的相关为

$$\begin{aligned}\mathrm{Corr}(F_i, F_j) &\to \mathrm{Corr}(\hat{b}^2_{yi.ai}, \hat{b}^2_{yj.aj}) \\ &= \mathrm{Corr}(\hat{b}_{yi.ai}, \hat{b}_{yj.aj})^2/2 \\ &= \gamma^2_{ai.aj}/2\end{aligned}$$

上式当 i 和 j 为侧连标记时等于0。

9.1.5 多元回归的性质

这里概括与QTL作图有关的多元回归分析的各种性质。Rodolphe和Lefort (1993) 也确定过这些性质。

假定对 n 个个体所组成的回交群体观测了数量性状和 t 个排序标记，并进一步假定按以下的线性回归模型进行数据分析

$$y_j = b_0 + \sum_{i=1}^{t} b_i x_{ji} + e_j \qquad (j=1, 2, \ldots, n) \qquad (5)$$

除在模型中包含了更多的标记外，模型变量和参数的定义与(1)相同。x_{ji} 是第 j 个体第 i 标记的类型。注意 b_i（也记为 $b_{yi.ai}$，s_i 为除第 i 标记外的所有其它标记所构成的集合）是在所有其它标记条件下表型 y 对第 i 标记的偏回归系数。上述分析已证实了这一分析具有下列性质：

性质 1　在多元回归分析公式(5)中，假定不同基因座位间 QTL的效应是可加的（即忽略上位性），则性状对某一标记的偏回归系数只取决于两个相邻标记所括区间内的 QTL，而与其它区间 QTL的效应无关。这一性质实质上是说根据偏回归系数可以构建条件(区间)检验。该检验满足了检验统计量与特定标记区间之外的所有其它 QTL的效应无关这一要求。

性质 2　在多元回归分析中以非连锁标记为条件，可以通过控制某些剩余的遗传变异来降低检验统计量的抽样方差，从而提高检验的统计效率。这意味着既使是非连锁标记也含有可用于提高检验的统计效率和遗传作图效率的有用信息。这些有用的信息尚未被其它 QTL作图法所利用。

性质 3　在多元回归分析中以连锁标记为条件将减少可能存在的多个连锁QTL对假设检验和参数估计的干扰，但也具有增加抽样方差的可能性。本句的第一部分重述了性质1，而第二部分则是说由于区间检验是一种条件检验，该检验可能丧失统计检验的效率（详见 Zeng 1993)。这概括了区间检验的优缺点，即利用区间检验时难以同时兼顾作图的精度和效率两个问题。这两个问题的有效平衡是实际的 QTL作图中要考虑的问题。

性质 4　在多元回归分析中性状值对两个标记的两个样本偏回归系数一般是独立的，除非这两个标记是相邻标记。这种独立性与两个区间位置之间检验统计量的相关性有关。

已经证实，对于区间检验，某一区间的检验统计量与另一区间的检验统计量一般是近似独立的 (Zeng 1993)。这一性质与在零假设下为覆盖整个基因组的总检验确定适当显著值有关。

9.1.6 回交群体以外的其它群体

以上提供了回交群体的理论分析。可以证实，与QTL作图有关的多元回归分析的类似结论和性质对于其它群体设计如F_1相互交配产生的F_2群体等也是成立的。例如，如果m个QTL不存在显性（即所有的$d_m=0$），则F_2群体的表型方差就是回交群体的有关方差的两倍。在F_2，每一标记有三种可能的基因型，当第h个体第i标记的基因型为P_1纯合体、杂合体或P_2纯合体时，统计模型中的x_{hi}取值分别为2、1和0。这样σ_i^2、σ_{ij}和σ_{yi}正好是回交群体相应方差的两倍。因而公式1和4的结果不变。公式(3)中的分母$\sigma^2_{i.ai}$以及分子$\sigma^2_{y.a}$的遗传部分也要乘以2，这样公式3中的$\hat{b}_{yi.ai}$的抽样方差将被降低。从基本上讲，上述4个性质不变。即使考虑显性，从定性上讲也是可用的，尽管在F_2存在显性时的分析要更为复杂，且为了处理显性变异，对于每一标记回归模型可能需要增加一个变量。但存在上位性时，性质1就不是完全真实的了。根据上位性的种类和程度，性质1在某些情况下可能近似成立。

9.2 复合区间作图

9.2.1 统计原理

直接将多元回归分析公式(5)用于QTL作图是不恰当的，因为偏回归系数很可能是有关QTL效应的有偏估计(Zeng 1993)。但是如上所述，对于设计一种准确而有效的QTL作图方法，多元回归分析具有几个明显的特征。根据多元回归分析的以上性质，将区间作图与多元回归相结合所发展的区间检验方法将能充分利用作图数据所能提供的信息。为简单起见，我们考虑回交群体的数据。该群体假定来自m个QTL和t个标记上不同等位基因固定的两个近交系。数据包括数量性状值和n个个体t个排序标记的观测结果。假定t个标记近似均匀地分布在基因组上。

假定要检验的QTL位于标记区间$(i,i+1)$上，可以利用标记i和$i+1$作为该区间内假定QTL的基因型的指标，将统计模型写作

$$y_j = b_0 + b^* x^*_j + \sum_{k \neq i, i+1} b_k x_{jk} + e_j \qquad (j=1, 2, \ldots, n) \qquad (6)$$

其中y_j是第j个体的性状值，b_0是模型的均值，b^*为在纯合体和杂合体之间表现出差异的QTL的效应，x^*_j是值为1或0的指标变量，两种值出现的概率取决于标记i和$i+1$的基因型以及假定QTL的检验位置(不考虑标记区间内的双交换重组，表9.1)。b_k为表型y对第k标记的偏回归系数，x_{jk}为第j个体第k标记的已知系数，当标记类型为纯合体时取值为1，为杂合体时取值为0。e_j为随机变量。模型中其它标记的求和取决于性质3两种因素的平衡(参见后面的讨论)。

假定e_j服从均值为0和方差为σ^2的等同和独立的正态分布，则似然函数为

$$L_1 = \prod_{j=1}^{n} [p_j(1) f_j(1) + p_j(0) f_j(0)] \tag{7}$$

其中，$p_j(1)$ 给出了 $x^*_j=1$ 的先验概率（表9.1），$p_j(0) = 1-p_j(1)$，$f_j(1)$和$f_j(0)$定义了均值分别为 $b_0+b^*+\Sigma_{k\neq i,i+1} b_k x_{jk}$ 和 $b_0+\Sigma_{k\neq i,i+1} b_k x_{jk}$、方差为 σ^2 的随机变量 y_j 的正态密度函数。将似然函数(7) 对各个参数求微分，并令其等于零，然后解方程，参数 b^*、b_k 和 σ^2 的最大似然 (ML)估计值就是下列方程的解：

$$\hat{b}^* = (Y-X\hat{B})'\hat{P}/\hat{c} \tag{8}$$

$$\hat{B} = (X'X)^{-1}X'(Y-\hat{P}\hat{b}^*) \tag{9}$$

$$\hat{\sigma}^2 = [(Y-X\hat{B})'(Y-X\hat{B})-\hat{c}\hat{b}^{*2}]/n \tag{10}$$

其中，Y是 y_j 的 $(n\times1)$ 的向量，$\hat{B}$ 是 b_k 的ML估计值（包括 b_0 但不包括 b^*）的 $((t-1)\times1)$ 向量，X 是 x_{jk} 的 $(n\times(t-1))$ 的矩阵，$\hat{P}$是确定 $x^*_j=1$的后验概率的ML 估计值、具有元素 $\hat{P}_j$ 的$(n\times1)$向量：

$$\hat{P}_j = p_j(1)\hat{f}_j(1)/[p_j(1)\hat{f}_j(1)+p_j(0)\hat{f}_j(0)] \tag{11}$$

且

$$\hat{c} = \sum_{j=1}^{n} \hat{P}_j \tag{12}$$

注意，如果将 p 处理为一个参数，则 p 的 ML估计就是下列方程的解

$$\hat{p} = \frac{\sum_{j=1}^{n2}(1-\hat{P}_j)+\sum_{j=1}^{n3}\hat{P}_j}{n_2+n_3} \tag{13}$$

其中，求和 $\sum_{j=1}^{n2}$ 和 $\sum_{j=1}^{n3}$ 是指属于表9.1中第2和3类群的个体的和。

表 9.1　指标变量 x^*的定义

类群	标记基因型		样本大小	x^*
	i	$i+1$		
1	+	+	$n1$	1
2	+	-	$n2$	1（概率为$1-p$） 0（概率为p）
3	-	+	$n3$	1（概率为p） 0（概率为$1-p$）
4	-	-	$n4$	0

+：标记基因型的纯合体；-：标记基因型的杂合体

在分析时，p 既可以处理为一个参数，也可以处理为一常量 $p=r_{iq}/r_{i(i+1)}$，其中 r_{ij} 为标记 i 与被检验QTL 的位置 q 之间的重组率，$r_{i(i+1)}$为标记 i和 $i+1$之间的重组 率。忽略标记区间内的双交换。

利用 ECM 算法 (Meng和Rubin 1993)，对上述式子进行迭代就可以获得这些估计值。迭代时可从初始估计值 $\hat{b}^*=0$ 开始，或从以 $x=p_j$ (1) 时所得 b^* 和 B 的最小二乘估值开始。每次迭代都包括一次E步骤 (公式 11) 和三次CM步骤(公式8、9和10)。Meng和Rubin (1993)证实当他们的条件 (3.6)满足时，迭代将收敛于最大似然估计值。与在M步骤同时使 $\hat{b}^*$ 和 $\hat{B}$ 最大化的全 EM 算法相比，这一算法的优点在于，并不需要对逆矩阵 $(X'X)^{-1}$ 进行更新修订，因而大大提高了数值评价的效率。

所要检验的假设为

H_0: $b^*=0$ 和 H_1: $b^*\neq 0$

在零假设下，似然函数为

$$L_0=\prod_{j=1}^{n} f_j(0)$$

ML估计值为

$$\hat{B}=(X'X)^{-1}X'Y$$

$$\hat{\sigma}^2=[(Y-X\hat{B})'(Y-X\hat{B})]/n$$

似然比(LR)检验统计量为

$$LR=-2\log(L_0/L_1)$$

$$=n(\log\hat{\sigma}^2-\log\hat{\sigma}^2)+\sum_{j=1}^{n2}[2\log(1-p)\exp(\hat{d}_j/2)+p]-\hat{P}_j\hat{d}_j]$$

$$+\sum_{j=1}^{n3}\{2\log[p\exp(\hat{d}_j/2)+(1-p)]-\hat{P}_j\hat{d}_j\} \qquad (14)$$

其中 $\hat{d}_j=[2\hat{b}^*(y_j-\hat{b}_0-\sum_{k\neq i,i+1}\hat{b}_kx_{jk})-\hat{b}^{*2}]/\hat{\sigma}^2$

当不存在上位性时，如果有标记可以将所考虑的QTL与其它连锁的QTL分开来，并且这些标记被拟合在模型中作为对照，则由这一方法所得QTL的位置 p 和效应 b^* 的估值将不受这些与之连锁的QTL的影响(性质1)。根据最大似然法的性质，这些估计值也将趋于渐近无偏(见表9.3)。但如果在QTL检验显著的条件下估计QTL的效应，则QTL效应的估计值将是有偏的 (表9.3)。因此将区间作图与多元回归分析相结合，就为将各个QTL分别进行检验和估计创造了条件。

但是，当QTL位于相邻的标记区间时，由于这两个相邻的标记用于区间作图，这些QTL就会干扰采用复合区间作图所进行的检验和估计。如果的确有两个主要的QTL存在于两个相邻的标记区间内，QTL的位置取决于区间及区间大小，则对于这两个区间和部分相邻的区间，似然图谱可能是极显著的，有时可能会表现出双峰特征。经区间检验，可能认为在两个相邻的区间存在两个QTL。此时，如有必要，可在这些区间拟合两个变量，对两个QTL和其它的标记一起进行检验，但为了检验假设，可能需要非常大的样本容量。

9.2.2 零假设下检验统计量的特性

与常规的简单区间作图法一样，本检验也可以对标记所覆盖的基因组上任何位置进行。因此这一检验为QTL 创造了系统的检测方法，并将多维检测多个QTL 的问题缩减为一维空间检测问题。由于这是一种（对多个位置的）多个检验，在实践中遇到的问题是对整个基因组进行时如何确定检验统计量的显著值。为了确定适当的显著值，需要了解在零假设下检验统计量的特性。由于对QTL 的检测是针对整个基因组进行的，因此我们不仅要知道某一特定检验区间而且要了解整个基因组上的检验统计量的特性。Lander和Botstein (1989) 讨论了简单区间作图法中检验统计量（LOD值）的显著值问题。但现在的复合区间作图方法检验零假设所需的检验统计量的临界值是不同的。其差别在于，多元回归的检验统计量在不同区间多少是独立的（性质3）。因此，为了获得对 M 个区间实施检验所需的总显著水平 α，对于每一区间检验可能会采用一种名义上的显著水平 α/M，这种情形与Lander和 Botstein (1989) 的松散型作图相对应。

表 9.2　在零假设下特定位置、特定标记区间和整个基因组上检验统计量的均值、方差和95%置信限

图距	n	每一位置平均			每一区间的平均最大值			极大值		
		均值	方差	95%	均值	方差	95%	均值	方差	95%
1	500	1.01	2.09	3.80	1.39	2.72	4.47	1.39	2.72	4.47
5	500	1.03	2.11	3.93	1.63	3.19	5.15	3.43	5.59	7.99
10	500	1.03	2.10	3.95	1.64	3.19	5.25	4.60	6.12	9.31
20	500	1.05	2.14	4.03	1.70	3.41	5.35	6.07	6.79	10.91
40	500	1.10	2.46	4.22	1.78	3.79	5.69	7.95	8.34	13.23
1	200	0.99	2.03	3.88	1.38	2.70	4.65	1.38	2.70	4.65
5	200	1.03	2.15	3.88	1.62	3.20	5.20	3.47	5.68	7.97
10	200	1.06	2.23	4.04	1.70	3.49	5.42	4.79	6.69	9.63
20	200	1.11	2.42	4.26	1.79	3.75	5.71	6.28	7.60	11.61
30	200	1.18	2.98	4.50	1.91	4.94	5.97	7.65	9.52	13.64
40	200	1.29	3.48	4.98	2.10	5.71	6.65	9.03	10.85	15.20

标记间的距离为10cM，模拟重复次数，样本大小为500时重复1000次，样本大小为200时重复2000次。

但是，在零假设下某一区间上LR 最大值的分布（即公式14以 p 作为参数的LR 的分布）是未知的。该分布在很大程度上取决于样本大小、模型所拟合的标记数目以及每一区间的大小。为了研究检验统计量的特性，在零假设下进行了一系列的模拟。尽管零假设是在有关的检验区间不存在QTL，但还是模拟了该基因组不存在QTL的情形以考察在基因组上检验统计量的特性。这在现在的模型中不会造成很大的差异，因为其它区间正在分离的QTL多少被模型中的条件标记所控制因而它们不会对被检验区间的假设检验造成很大的影响（参见Lander和 Botstein1989）。模拟了两种极端情形，连锁和不连锁。在连锁的情

形下，在一条染色体上存在 M 个均匀分布的标记区间（含 $M+1$ 个标记）；在不连锁的情形下，在不同染色体上存在 M 个均匀分布的标记区间（含 $2M$ 个标记）。对于每一次重复，都进行含 $M+1$（或 $2M$）个标记（每一标记区间的重组率为 r）和数量性状（简单地模拟为正态随机变量）的 n 个回交个体的模拟。根据上述方法，沿基因组每隔1cM区间位置按(14)式计算出LR检验统计量。

结果如图9.1和表9.2所示。表9.2给出了检验统计量LR在零假设下单个检验位置（即(14)式的LR取固定的 p 值）、单个区间LR的最大值（即(14)式的LR取 p 为参数）及基因组上LR的总体最大值的均值、方差和95%置信限。位置与位置之间、区间与区间之间LR的统计数一般存在极小的差异，因此只提供了所有检验位置上和所有标记区间上LR的平均统计量（均值、方差和95%置信限）。图9.1给出了样本容量（n）为500和200时LR的总体最大值的95%置信限对区间数的作图。

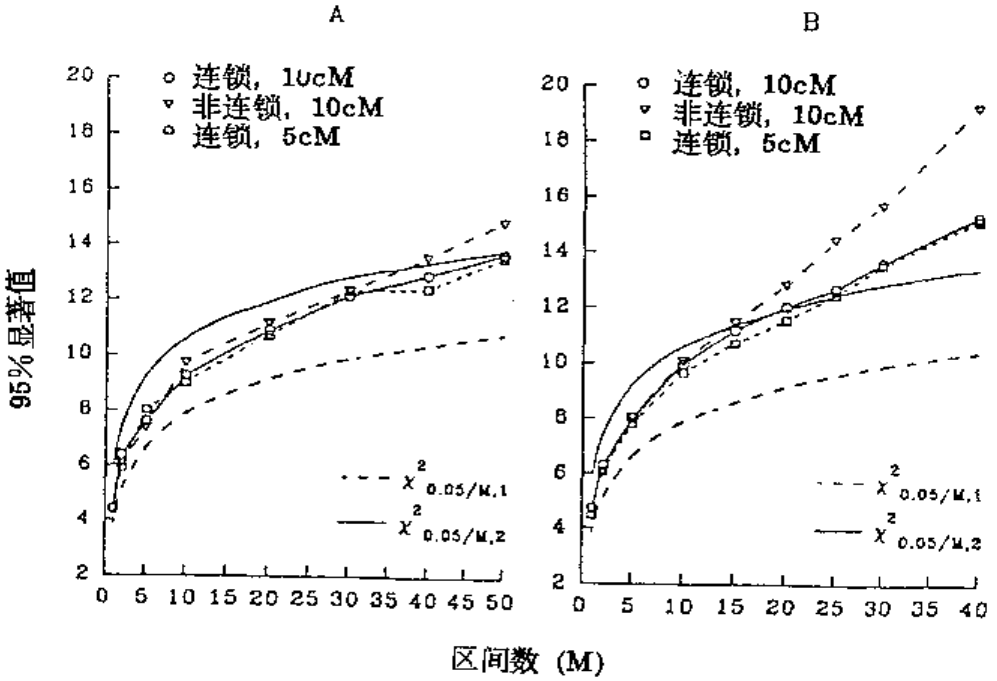


图 9.1 零假设下检验统计量极大值的95%置信限和区间数的关系

$\chi^2_{0.05/M,1}$ 和 $\chi^2_{0.05/M,2}$ 列于图上作为参照

图A的样本大小为200，重复2000次；图B的样本大小为500，重复1000次

非连锁：所有标记区间均为10cM，所有标记区间位于不同染色体

连锁：标记区间为10cM，所有标记区间位于一个染色体

连锁：标记区间为5cM，所有标记区间位于一个染色体

这些结果有力地说明，当样本容量较大、模型中所拟合的标记数相对较少时，固定检验位置（即对于固定的 p）的LR统计量近似地服从自由度为1（具有均值为1、方差为2、95%显著值为3.84）的 χ^2 分布，正如本来可以预期的一样（但参见Goffinet等1992的特例）。当样本容量相对较小、模型中所拟合的标记数相对较多时，所观察的LR偏离 χ^2_1 分布，这最有可能是因为缓慢地收敛于 χ^2_1 分布。

表9.3 QTL位置和效应的参数及100次模拟所得估计值的统计量

染色体		第 1 染色体			第 2 染色体		
参数	位置(cM)	16	48	108	3	43	77
	效应	0.42	0.75	0.58	1.02	-1.23	-1.26
	模型 I						
	效率[a]		0.22		0.87	0.83	0.80
	位置(cM)[b]		49.6 (6.4)		3.4 (2.9)	42.1 (3.4)	76.8 (2.9)
统	位置(cM)[c]		50.0 (5.2)		3.3 (2.4)	41.8 (3.0)	76.8 (2.6)
	效应[b]		0.83(0.26)		1.02(0.24)	-1.24(0.23)	-1.24(0.25)
	效应[c]		1.14(0.12)		1.08(0.20)	-1.29(0.21)	-1.32(0.20)
计	模型 II						
	效率[a]		1.00			1.00	
	位置(cM)[b]		48.2 (13.3)			71.9 (6.7)	
量	效应[b]		1.15(0.13)			-1.67(0.15)	
	模型 III						
	效率[a]		1.00			1.00	
	位置(cM)[b]		53.3 (19.2)			71.3 (8.9)	
	效应[b]		1.20(0.19)			-1.70(0.19)	

染色体		第 3 染色体			第 4 染色体
参数	位置(cM)	33	68	129	26
	效应	-0.46	1.61	0.88	0.74
	模型 I				
	效率[a]		0.99	0.58	0.17
	位置(cM)[b]		68.8 (2.2)	130.3 (2.5)	27.1 (6.2)
统	位置(cM)[c]		68.9 (2.1)	130.0 (2.2)	27.9 (4.5)
	效应[b]		1.62(0.26)	0.94(0.23)	0.78(0.26)
	效应[c]		1.63(0.24)	1.15(0.16)	1.12(0.19)
计	模型 II				
	效率[a]		1.00	1.00	1.00
	位置(cM)[b]		70.4 (2.3)	121.5 (11.0)	27.0 (5.4)
量	效应[b]		1.65(0.13)	1.33(0.14)	0.74(0.13)
	模型 III				
	效率[a]		1.00	1.00	1.00
	位置(cM)[b]		70.3 (2.6)	121.9 (10.9)	27.6 (7.1)
	效应[b]		1.67(0.18)	1.34(0.19)	0.76(0.20)

[a] 根据100次重复模拟结果在指定区域检测QTL的效率的经验估计

[b] 根据全部100次重复模拟所得估值的均值及标准差(括弧内数据)

[c] 由在指定区域检测显著的模拟所得估值的均值和标准差(括弧内数据)

在所考虑的情形下标记区间上LR的最大值一般落在 χ^2_1 和 χ^2_2（均值为1，方差为4，95%置信限为5.99）的分布之间（n=200且M=40的情形例外，此时检验的自由度显著减少）。分布取决于样本大小、模型中所拟合的标记数目以及区间的大小。区间之间LR最大值的相关一般极小（且为正）（性质4）。（相邻标记区间的平均相关一般小于0.3，非相邻标记区间的平均相关基本接近于0）。与其它因素如样本大小、模型中所拟合的标记数及被检验的区间数相比，标记区间的大小（以图9.1 中的10cM 区间和5cM 区间相比较）的影响较小。因此，当样本容量较大而模型中所拟合的标记数不是很大时，$\chi^2_{\alpha/M,2}$ 可用作近似的(1-α)100% 显著值以对模型(6)基因组上 M 个区间进行总体检验（图9.1A）。（尽管图9.1只提供了95%的显著值，其它如90%的显著值也表现类似的模式）。

但是在实际上，作图数据的样本容量并不是很大，标记也不太可能在染色体上均匀分布。这时如在模型中拟合太多的标记，即使它们与被检验的区间不是连锁的，也可能并不是恰当和理想的，因为在模型中拟合太多的标记实质上增加了检验统计量的显著值（例如，比较图9.1B中连锁的(M+1)个标记和非连锁的(2M)个标记），因而降低了检验的效率。在作图数据中许多标记通常聚集在某些区域。如果这样，更为恰当的方式是从基因组上选择使之分布较为均匀的那些标记，或事先已经（如利用逐步回归）鉴别的、可以解释基因组大部分遗传变异的那些标记拟合在模型中，以作为对照。其它的标记仍用于有关区间的QTL检验。对于许多数据，上述指标可以用作一个参考（并不一定要 α=0.05）。必要时可用计算机模拟来确定特定数据的检验所需的适宜临界值。

9.3 模拟分析

9.3.1 方法

为说明本方法的性质、应用及优缺点，Zeng (1994) 进行了QTL作图的模拟研究。模拟了4个染色体，每个含16个标记，构成15个10cM区间的回交群体。性状受10个QTL的影响，其位置和效应列于表9.3，图谱列于图9.2。所有QTL 共解释回交群体中表型变异的70%。共模拟了300个回交个体。一个个体的性状值由两部分组成，一是该个体所具有的QTL的效应之和，二是均值为0、方差使群体遗传率为0.7的正态随机（环境）变量。根据以上定义的连锁图，模拟每一个体的标记类型和QTL类型。

这一数据可用于分析拟合大量的模型。但作为说明问题，根据以上所概括的程序，只拟合了以下三种模型：

模型 Ⅰ：利用拟合在模型中的所有标记进行复合区间作图（图9.2中的实线）

模型 Ⅱ：利用侧连标记和拟合在模型中的所有非连锁标记进行半复合区间作图（图9.2中的长虚线）

模型 Ⅲ：仅利用模型中所拟合的侧连标记进行简单的区间作图（图9.2 中的短虚线）

图9.2中的点线给出了通过5000次重复模拟所得的检验模型 Ⅰ 所需的总体5% 临界值(16.5)，该值略高于 $\chi^2_{5.05/60,2}=\chi^2_{5.00083,2}=14.2$。模型 Ⅱ 和 Ⅲ 的总体5% 临界值要小一些，分别为12.6 和10.7。 说明QTL存在区域的似然图谱顶峰处QTL的位置和效应的点估计值列于表9.3。 图9.2和表9.2都显示了单个重复模拟的作图结果。

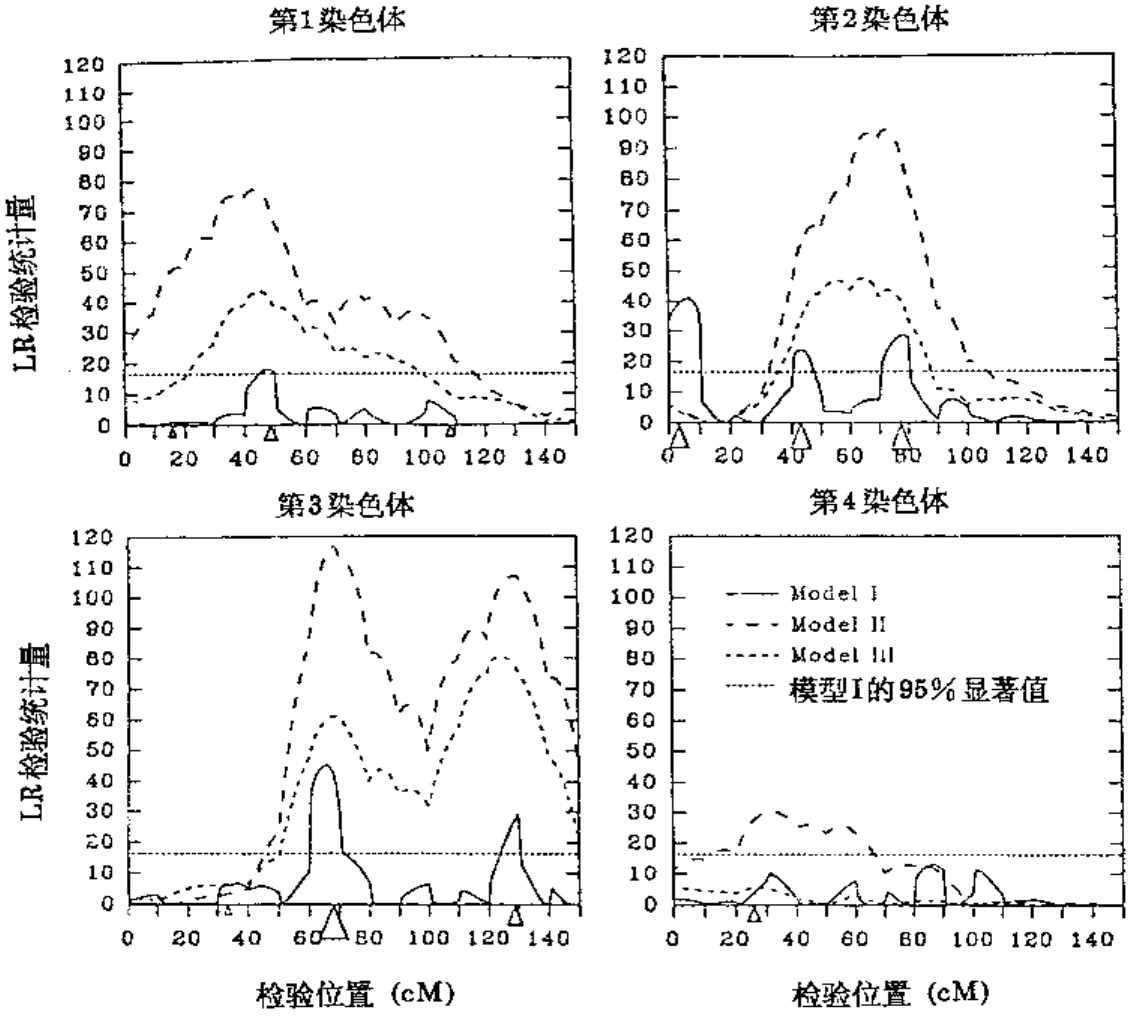


图 9.2　回交群体QTL作图的模拟结果

每隔1cM计算似然比检验统计量以描绘似然图谱，每染色体长150cM，每10cM分布一个标记，10个模拟 QTL的位置用三角形标出。三角形的大小与QTL的效应成比例，粗线三角形表示QTL 的效应为正，细线三角形表示 QTL的效应为负。三种曲线表示三种不同的模型：模型 I （实线）为复合区间作图；模型 II （长虚线）为半复合区间作图；模型 III（短虚线）为简单区间作图。 点线为模型 I 的5%模拟临界值，16.5。

为了说明结果的一般形式，根据同样的参数和分析方法，重复模拟了100次。表9.3给出了图2所示QTL位置和效应估计值的均值和标准差。对于每次重复，都在有关区域附近的似然图谱顶峰处进行参数估计。尽管图2只是根据一次重复的结果所作的分析，但得到100次重复作图结果的证实。 但是，也有少数重复，利用模型 II 和 III 作图时在染色体3上只检测到一个而不是两个QTL。这时，为了使结果可以与其它重复比较，仍将其解释为检测到

两个QTL，并根据似然图谱的有关的区域性最大值来确定QTL位置和效应的估计值。表9.3还提供了这种检验的效率的经验估计值(即在有关区域显著的重复数与总重复数之比)。

9.3.2 结果

无论模型 II 或 III 都不能控制位于被检验染色体其它区间上的QTL对特定染色体位置的QTL 检验的影响。但是，模型 II 可以从模型剩余方差中有效地去除由于非连锁染色体上QTL 分离所产生的大多数变异，因而在该模型下大多数区间的LR 检验统计量显著高于另外两种模型的(性质2)。模型 II 检验标记-QTL 连锁的统计效率最高。但是，由于该模型(以及模型 III)下的检验不是区间检验，因而并不一定具有较高的概率来准确地定位单个QTL，因为QTL 效应和位置的估计值可能受到其它连锁QTL的影响。只有模型 I 提供了一种区间检验，检验统计量与被检验区间外的 QTL和两个彼邻区间的效应是独立的。这是区间检验的一个优点。模型 I 也可以从模型剩余方差中有效地去除分离QTL所产生的大多数变异。但是，由于模型 I 的检验是一种条件检验(性质3，见Zeng 1993关于理论背景的详细讨论)，故许多区间的检验统计量仍然较模型 II 和 III 为小。这是区间检验的一个缺点。

但无论如何，根据似然图谱 (LR 检验统计量)显著峰的相对准确位置和似然图谱的斜率来判断，模型 I 正确而精确地鉴别了图9.2中的6个最大的QTL。以附近其它的标记为条件进行检验，可以提高检验统计量对 QTL所在位置的敏感性，使QTL的位置效应明显地表现在被检区间的小区域内，从而实现精确作图。这是区间检验的另一个优点。

在第3染色体上模型 II 和 III 给出了似然图谱的两个显著峰，正确地指明了存在两个主要的 QTL (第3个 QTL 效应较小)，这是因为这两个主要的QTL 相隔了足够的遗传图距。但是在第2染色体上，模型 II 和 III 似乎在一个错误的位置指出存在单个 QTL，这是由于在相对紧密的范围内存在具有相当 (或相反) 效应的三个 QTL。如果发生这样的情况，我们就会被错误的模型所欺骗了。当然，可以采用 Lander和 Botstein (1989)所建议的方法对这些数据同步拟合和分析两个或三个 QTL，但两个 QTL／一个QTL 和两个 QTL／三个QTL 的统计检验是不容易的。在实际应用时，在某一染色体上 QTL 的数目(即所检验的真实遗传模型)是未知的，而且非常容易拟合和检验一个错误的模型。既使在模型 II 和 III 下可以鉴别 QTL 的正确区域的场合，其似然图谱也趋向于更为扁平 (即Lander和 Botstein (1989)所定义的支撑区间更宽)，作图精度仍然相对较低(表9.3)。

似乎只有模型 I 才能给出QTL 位置和效应的相对无偏的估计(表9.3)。但是，如上所述，模型 II 在鉴别 QTL 的可能区域方面具有最高的统计效率，而模型 I 由于拟合了紧密连锁的标记而使其鉴别 QTL 的统计效率相对较低(性质3)。只有模型 II 指出第4染色体上存在QTL。在实践上，为了最大可能的检测到QTL 同时控制作图的精度(即权衡统计检验上的 I 类和 II 类错误)，在数据有限时，可能必须利用模型 I 和 II 的某种形式的组合(即在模型中消去或插进一些连锁的标记作为对照)。作为例子，可以图2所给第1染色体的模拟结果作进一步分析。在该染色体上，模型 II 和模型 III 都显示出似然图谱有一个主峰，表明在第5区间有一个 QTL。但模型 II 显示在70cM 和110cM 之间还存在第二个QTL，而在模型 I 下不能证实其显著地存在。为检验这一假设，在0cM 和50cM 之间增加6个标记以控制背景，采用模型 II 在60cM 和150cM 之间的区域进行检验，应该可以消除第5区间上已定位的

QTL(以及位于50cM左侧的其它可能的 QTL)对检验和估计的影响。的确，检验有力的说明存在另一个QTL，在100cM 处的效应估计为0.52(图9.3)。

关于QTL效应估计值的准确性，必须区分影响估计值相对准确性的两个问题，即偏倚和抽样误差。与模型 II 和 III 相比，在模型 I 下所得 QTL 效应估计值的偏倚要小些，这是因为模型 I 的估计不受其它区间 QTL的影响(性质1，表9.3)。但模型 I 的抽样方差大于模型 II(性质3，表 9.3)。如果已知在某染色体上存在一个以上的QTL，则模型 II 就具有较高的检验 QTL的统计效率，并且QTL效应估计具有较小的抽样方差(例如表9.3的染色体4)。否则，利用模型 II(以及模型 III)所进行的检验和估计就可能导致 QTL 定位的错误。

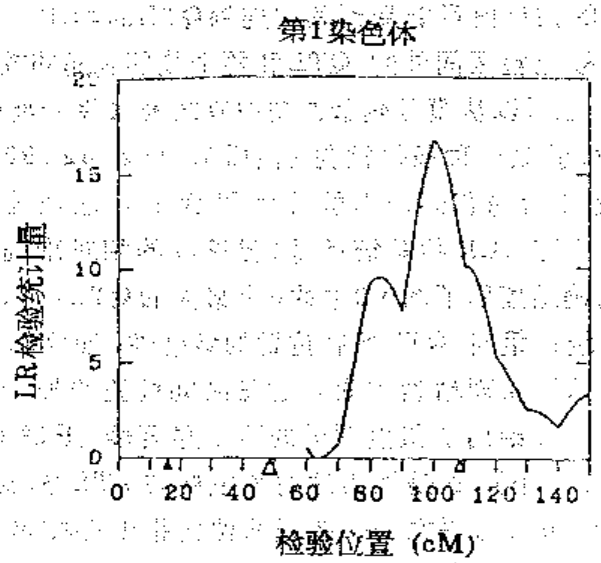


图 9.3　图 9.2 所示第 1 染色体 60cM 和 150cM 之间区域上的 QTL 作图结果的再分析
分析时拟合了位于第 1 染色体 0～50cM 区间的 6 个标记和所有其它非连锁标记

必须承认，为了控制遗传背景，有可能在模型 II 和 III 中利用了太多的标记。一般来说，标记在 QTL 作图中起个三个作用，这些作用因所检验的染色体位置而不同。标记可用于构建区间作图所需的指标变量 x^*，可以提供区间检验的边界，可在检验和估计中用于控制剩余遗传变异。对于第三种作用，只有当标记与 QTL 相连锁时才是有效的。对于每一染色体，可能只有少数标记与 QTL相连锁(当然不知道在染色体上有多少QTL)。如果在模型中只拟合与QTL 紧密连锁的一些标记，则其它的标记都是不能提供信息的 (性质1)，在模型中拟合也是多余的。拟合多余的标记可能降低统计检验的效率，并增加估计值的抽样方差，当样本较小时尤其如此。在实践上，可能需要采用逐步回归之类的方法先从染色体上选择一些标记，并在区间作图时利用这些标记来控制遗传背景。

对于QTL 作图，无偏性和准确性可能比作图技术的其它问题如计算的容易等更为重要。由于数量性状变异一般受到多个基因的影响，在设计作图方法时，必须考虑多个基因的可能效应。单个基因效应的有效控制是提高作图精度的关键。在作图时应尽每一个努力来保持作图的准确性和质量，同时最大限度的提高找到更多QTL 的机会。

事实上，Zeng (1994)倡导了一种新的QTL作图框架。这一方法的关键特征是在QTL检验和作图时试图将单个QTL的效应区分开来的区间检验思想。根据数据和起作用的遗传机制，可显著提高作图的精度。对于遗传率较高的性状，由于大多数的遗传变异可以被控制和从以标记为条件的模型的剩余变异中去除，这一方法更为适合。对于遗传率较低的性状，在模型中拟合多个标记模型的所得可能无法弥补因丧失检测QTL的统计效率之所失，因为如果在模型中为区间检验拟合紧密连锁的标记，上述两个方面都会使估值的抽样方差增加(性质3)。这样也就降低了检验统计量，增加了检验的临界值(这是由于两者都使独立检验区间的数目增加、检验自由度减少)。这时，首要的任务可能是提高基因组上检测QTL的机会。在实践上，最好的策略可能是用多个模型来拟合数据以挑选出可以平衡 I 类和 II 类错误的恰当模型。

Zeng (1994) 只分析了回交群体设计。象F_2等其它的群体设计及显性遗传模型很容易移置进去。但是这里忽略了上位性。如果QTL存在上位性，本作图法可能仍然是有偏的。在 QTL作图分析上位性所遇到的问题是存在多种类型的遗传上位性，并且尚未很好地定义起作用的上位性遗传参数。在原理上，如果遗传上位性的类型可以鉴别，就可以在QTL作图模型中拟合上位性效应(例如 Haley 和 Knott 1992)。

Jansen (1992)、Haley 和 Knott (1992)、Martinez 和 Curnow (1992)、Moreno-Gonzalez (1992)、Jansen (1992,1993) 已采用各种方式将回归分析用于QTL作图。尽管 Haley 和 Knott (1992) 以及 Martinez和Curnow (1992) 都曾建议利用二元回归分析沿着被检测的染色体在二维检验空间检测以定位两个QTL，但只是简单地将多元回归用于近似地获得简单区间作图方法。Moreno-Gonzalez (1992) 建议在QTL作图中利用逐步回归方法来拟合多个标记(将多个QTL随机地假定其存在于每一标记区间的中部)，显然这种方法是非常随机的和不精确的(Zeng 1994)。Jansen (1992) 描述了以区间作图为骨架的混合模型(Mixture model)分析法，但他给出了一个以第三个标记为线性模型中的协变量作为对照的模拟例子，这与现在所倡导的方法有些相似。最近，Jansen (1993) 也倡导了一种将多元回归分析与区间作图相结合的 QTL 作图方法，但与复合区间作图方法即有相似也有不同。很显然，在定位多个 QTL 的概念和方法上均存在重要的差异。复合作图法采用的方法是区间作图，强调尽可能地控制作图的精度。而Jansen利用同时检验一个染色体上的多个标记的方法来检测可能存在的多个 QTL，Zeng(1994) 认为这对于多个 QTL 的作图可能是很不精确的，而且Jansen没有讨论和分析有关的理论和统计性质。

概括地说，复合区间作图方法有四个优点。(1)将检测限于一次一个区间，将(多个QTL 的)多维检测问题简化为一维检测问题，并且单个 QTL的效应和位置的估值可能是渐近无偏的。(2) 在检验中以连锁标记为条件，极大地提高了 QTL 作图的精度。(3)该方法同步利用所有信息进行推断，并且应比其它方法对 QTL 作图更富有信息、更为有效。(4)仍然可以利用 QTL 似然图谱来表示整个基因组上每一点上QTL 的证据强度，从而保留了区间作图的特性。一个完整的连锁图谱不仅可以用于为标记所覆盖的基因组上任何地方的QTL 确定检验的位置(区间作图)，而且可用于提供检验的边界条件，同时控制基因组其余位置的剩余遗传变异 (区间检验)。因此上述优点是显而易见的。

第 10 章　数量性状的选择与育种

数千年来，人们通过选择高产、食味好或其它感兴趣的植株来培育新品种。不管是古人早期选育的作物品种，还是后来经当代育种家改良的作物品种，都是形成今天复杂的、高生产力的经济基础。现代遗传改良计划一般是针对现有某品种的一个或多个有缺陷的性状。而该品种的遗传改良需要鉴别与缺陷性状的理想表现型有关的等位基因，并将这些基因导入到该品种之中。当有关等位基因数量较少且对表型有较大的效应时，对遗传材料中进行简单的表型筛选就可鉴别理想的等位基因，然后通过经典的育种程序（杂交、回交、自交和选择）很容易将其导入其它品种之中。育种家根据表现型和基因型之间的明确关系，就可以检测出有关群体中的理想等位基因。

但是，经典的育种程序，其遗传操纵仅限于该品种所属生物种的范围内。通过整基因组杂交、独立分配和重组来产生有利重组体，并从许多分离群体中将其鉴定出来。即使对于具有主要效应且表型效应截然可分的等位基因，也需要许多世代才能完成。而当有利基因与不利基因座位紧密连锁时，这项任务几乎是不可能完成的。

另一方面，尽管包括抗病性等在内的许多重要性状是由具有主要表型效应的基因决定的，但象产量、品质，对逆境的水平抗性等重要经济性状都属于数量性状。影响这些性状的遗传差异是由多个基因座位决定的，每个座位对最后表型只起较小的正向或负向贡献。这类数量性状的表型表达受许多基因座位遗传变异的影响。此外，由于遗传和环境因素都以相同的方式对性状值起增效或减效的作用，因此仅仅根据性状的类型分布一般无法来区别遗传因子和环境因子对变异的作用。

尽管存在这些限制，借助于特定的试验设计和数据分析方法(生统或数量遗传学，Mather和Jinks 1982; Falconer 1981），可在某种程度上了解数量性状的遗传结构，并用于指导遗传改良计划。但是，在许多情况下，无法区别影响重要性状的各个基因座位，并在育种计划中单独操纵它们，从而限制了经典遗传改良方法的效率。同样，数量性状基因难以寻踪也给其基因克隆及其在遗传工程计划中的利用造成巨大问题。因此，目前在植物遗传改良中可供采用的两种方法（经典育种和遗传工程）处理数量性状的能力是有限的。分子标记为把 QTL转化为经典育种计划中易于操纵或遗传工程中可以克隆的孟德尔或准孟德尔单位提供了工具，从而填补了经典育种方法和遗传工程技术之间在改良植物方面存在的鸿沟。以分子标记及其遗传图谱为基础的技术可以帮助育种学家更好地了解数量性状的遗传基础，并为遗传育种家定位和操纵与数量性状表达有关的单个基因座位提供有效的途径。特定性状QTL 的鉴别将为上位性，多效性及杂种优势遗传基础的研究提供更有力的工具。已定位遗传标记的应用将大大加快植物改良和选择的进展。本章主要讨论分子标记及其遗传图谱在植物数量性状选择和育种中的应用。

10.1 回交导入有利基因

今天的作物品种是长期以来直接选择和间接选择的结果。随着有关的品种或种中理想性状的导入，使品种逐步得到改良。这一过程需用带理想性状的材料（供体亲本）与待改良的品种（轮回亲本）进行杂交。杂交得到的分离群体是亲本染色体片段的嵌合体，不仅带有供体亲本的理想性状，而且也带一些不利的性状。在导入供体亲本的基因之后，回交育种技术可使轮回亲本的基因型变得更理想。回交时，供体亲本与轮回亲本的杂种再与轮回亲本杂交，在后代中选择含目标性状的植株再与轮回亲本杂交，不断重复这一过程。几个循环后，就可得到遗传背景几乎与轮回亲本相同、但已导入目标基因的植株。

回交育种中一个长期未决的问题是有利基因与不利基因的连锁。即使回交20代后，还能发现相当大（10cM）的与目标基因连锁的供体染色体片段（Stam和Zeven 1981）。大多数植物基因组中10cM的DNA足够包含几百个基因。因此，回交不仅转移了目标基因，而且也转移了与目标基因连锁的其它基因。这一现象称连锁累赘（Zeven 等1983），它常使性状改良后的新品种与原来的目标不一致。根据分子标记进行选择可避免或至少显著地减轻连锁累赘的程度。传统的回交过程中，许多世代的回交之后连锁片段还很大，这并非是由于该区域未发生重组，而是我们缺乏有效的方法去鉴定重组个体，通常凭运气偶尔选择到这种供体片段较小的重组个体。而采用高度饱和的分子图谱有可能直接选到在目标基因附近发生重组的个体。根据过去理论推算和RFLP实验研究的证实，即使经过几代的回交，目标基因两侧仍包含很长的供体亲本片段。运用与目标基因紧密连锁的分子标记，则可对该区域的重组体进行筛选。这样，可能在较少的的回交世代中缩短该片段的长度。

10.1.1 标记辅助技术的特点

利用分子标记技术有希望克服回交育种的局限性。如果要转移的基因与分子标记紧密连锁，就可在苗期（性状表达前）从植物分离群体中筛选含目标基因的植株。由于分子标记既可标记主基因，也可标记QTL。因此，利用标记辅助的回交选择不受转移性状类型的限制，如在番茄上，已经用RFLP标记了影响果重、pH值和可溶性固形物含量的微效基因（Paterson 等1988；Osborn等1987；Tanksley和Hewitt 1988），这就有希望用回交方法快速地在品种之间转移这些基因。没有连锁的分子标记，要在育种实践中监测基因的流向是极为困难和费时的。

Tanksley和Rick（1980）和Tanksley（1983）讨论了利用同工酶标记来加速将主基因控制的性状从外源野生种质渐渗导入栽培品种的问题。这一过程中，普遍的问题是要在保留来自外源供体目标基因的同时，尽快地消除供体基因组，并用轮回亲本（受体）基因组取而代之。Tanksley和Rick（1980）和Tanksley（1983）建议将标记辅助的回交技术用于野生种和栽培种之间的基因渐渗转移。但是如果利用RFLP差异性地标记染色体，这一技术也可以用于同一物种不同品种间的基因渐渗转移（Nienhuis 等1986）。

经典方法转育隐性目标基因时要在每一或每隔二个回交世代后进行世代测交，以确认目标基因的存在。而借助多数呈共显性遗传的标记选择，转育工作可通过不间断的回交完

成。所以，借助标记的选择是加速抗性育种工作的一种手段，在世代间期长的植物中尤其如此。

10.1.2 所需回交代数

来自供体亲本的不良基因经轮回亲本的重复回交可逐渐消除。Allard (1960) 指出在多数情况下6个回交世代就足够了，但远缘杂交需要更多的回交世代。尽管重复回交通常是转移单基因行之有效的方法，但它需要花费很长时间，对世代周期长的物种尤其如此。假定在整个基因组中散布着大量标记，最好是每个染色体或染色体臂至少有一个标记。第一次回交后代或任一早期分离后代中，携有目标基因的植株可在苗期根据它们的标记基因型进行筛选和淘汰。轮回亲本标记基因纯合比例最高的个体保留下来作为下一世代的亲本。根据被筛选的标记个数，一次回交所选个体中轮回亲本基因组所占比例与未作标记选择下三次回交的比例相当。然而，值得强调的是回复到轮回亲本基因型的有效选择需要筛选数百乃至数千个个体。从经济方面考虑，只有在标记分析廉价和极需节省时间时，这种方法才较适合。

对于借助于目标基因与单个标记的连锁进行选择，我们假定每一回交世代产生 30 个植株，根据轮回亲体基因组的百分率从中选择最佳植株作为下一代回交的亲本。用20次独立的模拟发现要恢复轮回亲本基因组，采用传统回交育种方法，平均需6.5±1.7代。根据标记辅助的回交育种计划，恢复到轮回亲本的基因组只需3代(参见表10.1)，即经过3次回交就足以消除外源遗传背景 (Young和 Tanksley 1989b)。

利用目标基因两侧的标记(双标记)进行选择性回交，可以更快地消除非轮回亲本的染色体片段。在回交一代，用与目标基因紧密连锁的分子标记来鉴定在目标基因一侧1cM处发生交换的个体。这些重组个体再与轮回亲本回交，在BC_2中用与目标基因紧密连锁的另一分子标记来选择在目标基因另一侧1cM处发生交换的重组个体。估计表明，在约150个回交植株中，至少有一株在目标基因一侧1cM处发生一次交换的机率为95%。分子标记能明确鉴别这些植株。在含300个植株的再次回交中，目标基因另一侧1cM处发生一次交换的概率为95%，产生的含目标基因片段的长度不大于2cM。通过2个回交世代，用分子标记进行选择就可达到这一步。如不借助于标记则平均需100代才能达到(图10.1)。选到某一染色体区域内理想重组植株的可能性显然取决于该区域内已知标记的数目和测定植株的数目。随着分子标记图谱越来越饱和，选择重组型个体的效率将会不断提高。图10.1A表示与导入基因不连锁的基因组区域恢复到轮回亲本基因型的速率。图示基因型是在单个BC1植株的各回交世代中随机选择一些植株通过计算机模拟得到的。图中只给出了番茄每对同源染色体中的一条（另一条同源染色体均来自轮回亲本）。深色区表示供体基因组片段，浅色区表示发生交换的片段；空白区表示轮回亲本的基因组片段。每一区间约2cM。图下的数字表示来自轮回亲本基因组的百分率。图10.1B表示回交过程中目标基因二侧的连锁累赘分析结果。其上图是传统回交育种，下图是根据分子标记选择目标基因附近染色体发生重组的植株。图下面标出了获得特定量连锁累赘所需的年数（以每代需0.5年计算）。

一旦建立起分子标记与QTL 的紧密连锁，也可以通过回交将有利等位基因导入到理想的栽培品种或亲本品系中。如果 M_1是与有利 QTL等位基因 Q_1连锁的标记等位基因，那么

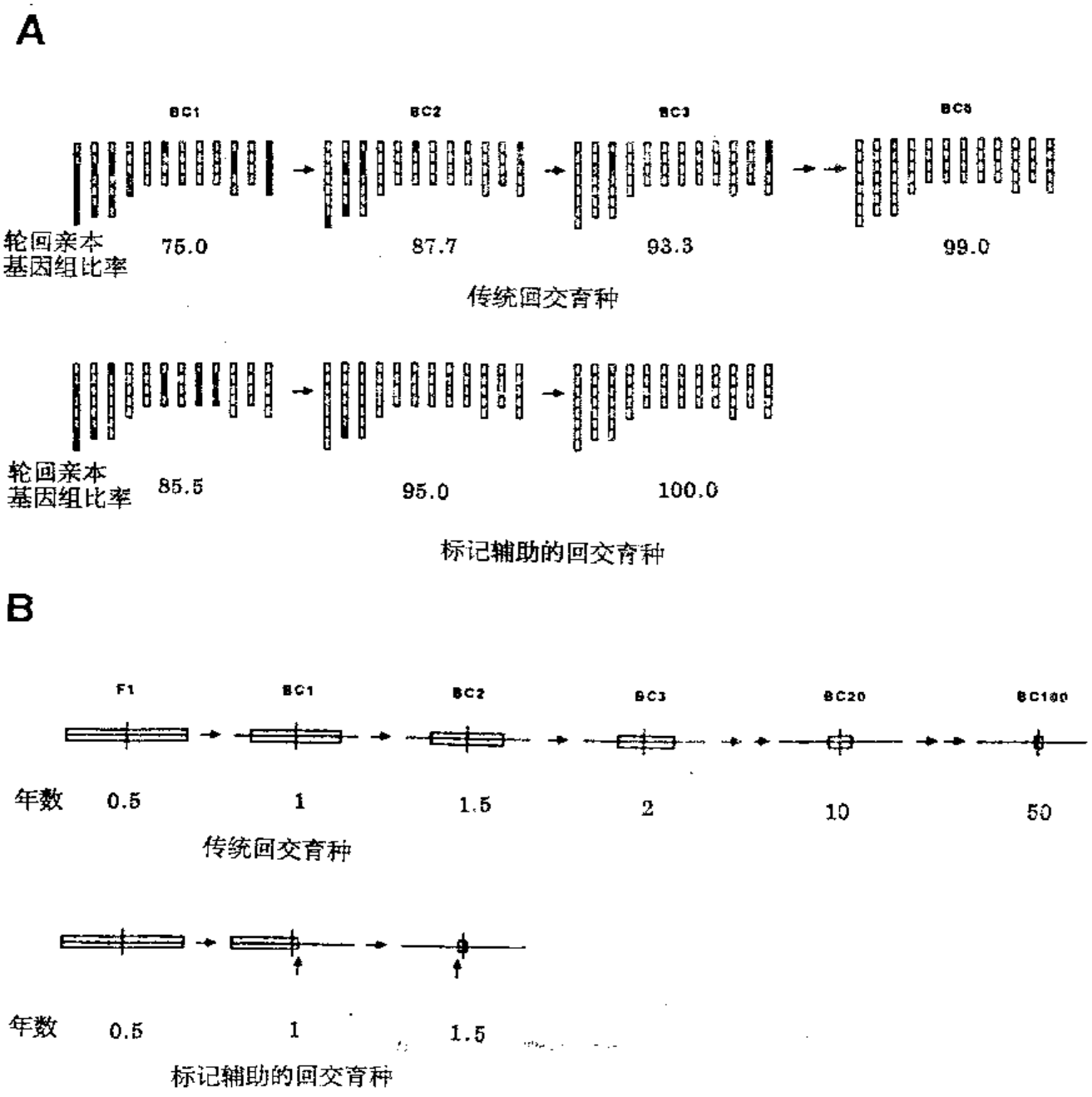


图10.1　传统回交育种与分子标记辅助的回交育种的比较

在回交过程中，就可以使 M_1M_2 杂合体的频率通过选择保持在0.5 的水平。但是，与 Q_1 等位基因相关联的 M_1等位基因的频率将因M_1和Q_1间的连锁而变化。如果 r=0.5，则带有 M_1的 BC_1个体中只有50%也是 Q_1。但是，如果 r=0.01，则带有M_1的 BC_1中99%为Q_1。回交次数越多，标记与 QTL之间的紧密连锁就越重要。当标记－QTL 之间的 r=0.01时，带有理想标记等位基因的 BC_3个体中，97%将带有理想的 QTL 等位基因。有趣的是，利用单一标记和双标记(括号标记)进行三代标记辅助的回交之后导入等位基因的频率分别为0.66和0.85，而不借助于标记经6次回交，导入等位基因的频率仅为0.01(表10.1)。

该回交程序对于少数基因座位控制的性状是恰当的。但随着处于分离状态的 QTL数目的增加，为了保证具有较高的概率在所有座位上获得有利的等位基因标记，所必需种植的回交个体数将随之增加。利用相邻的标记可以将QTL与成对标记更为紧密地连锁在一起，但它降低了在单个个体中发现所有有利标记等位基因的概率(Dudley 1993)。此外，将有更大比例的供体亲本与标记相连锁。如果该供体亲本不理想，就更难于在回交过程中去除掉。正在被转移的QTL数目越大，被转移到回交品系中且与标记座位相连锁的供体亲本基因组就越多。

回交育种与分子标记技术相结合，可以快速改良现有品种的某一特定性状。能用分子标记检测的基因（主基因或QTL）都能迅速、有效地被转移到另一个品种中去。一旦鉴别出足够数目的基因，就有可能对栽培品种的重要农艺性状如成熟期，株高和品质因素等进行精细的改良。用多个探针进行筛选几乎不需要付出额外的努力，故可考虑同时把多个基因导入到一个品种中。

表10.1　标记辅助的基因导入过程中特定回交世代的有利等位基因频率

回交代数	选择有利等位基因（有利等位基因的频率）			淘汰其余的外源基因组（受体基因组回复比例）	
	无标记选择	单一标记选择[1]	括号标记选择[2]	无标记选择	标记完全覆盖[3]
1	0.25	0.81	0.92	0.75	0.85
2	0.12	0.73	0.88	0.88	0.99
3	0.06	0.66	0.85	0.94	1.00
4	0.03	0.59	0.82	0.97	1.00
5	0.02	0.53	0.78	0.98	1.00
6	0.01	0.48	0.75	0.99	1.00

[1]标记等位基因与有利等位基因之间的重组率为10%
[2]两个括号标记之间的重组率为40%
[3]假定每一染色体上有2个标记

10.1.3　回交育种计划的设计

迄今，对适合于标记辅助选择的最佳育种设计知之甚少。特别令人感兴趣的是在每个世代中应检测和保留的个体数，而且这些问题与连锁程度以及是否采用单个或双标记基因有关。图10.2表示了在轮回杂交中，对目标主基因借助标记选择时所涉及的育种步骤和决策。产生 BC_1世代后的第一步工作是对各个体进行标记分析(第1步)，保留理想标记等位基因杂合的所有基因型（在单标记时为单杂合，双标记时为双杂合），然后与轮回亲本杂交（第2步)。由此产生的具有不同个体数(L) 的 BC2家系按下列途径之一对标记基因型进行筛选(第3步)。选择 I 为每次顺序检测一个家系中的一个个体，直到获得一个期望的标记基因型个体或检测了 k 个植株后仍未找到为止。选择 II 为同时筛选每个家系中 $L=$

k 个个体，若这些个体中无期望的标记基因型，则淘汰该家系，否则将该家系中的理想标记基因型个体用于继续回交（第4步）。第3和第4步可持续几个世代。在结束回交时进行标记分析（第5步），家系中带有理想标记基因型的个体将根据直接测验或世代测验评价其目标基因的有无（第6步）。育种工作的最终目标是在 BC_b 世代中至少获得 R 个衍生于不同 BC_1 个体的目标基因携带者，这些基因型可再用作回交或自交的亲本，在对其进行农艺性状的评价之后，作为新品种或杂交种、综合品种和多系品种的品系来推广。

10.2 近交选择

总体经济价值大致相等的两个栽培品种可能在其各自的优点和弱点上存在差异，或者在它们大致相等的特定性状上存在精细遗传结构的差异。这时，我们可以培育出结合两个品种优点或者在某一特定性状上超亲的重组近交系。这一系谱育种计划就是将两个栽培品种杂交产生 F_2，然后自交，跟踪和评价每一个家系，直到 $F_6 \sim F_8$ 代，形成数百个重组近交系并对其进行选择。随着自交过程的继续以及每一家系的特性逐渐固定，首先按株行然后按小区种植对每一世代保留的家系进行鉴定，最后在大田进行大规模试验更为精确地鉴定农艺性状。显然，这要求在每一世代大大减少鉴定的家系数，有时可能要借助于选择压以保留优良的重组自交系。 也就是说，在 F_2 代可能包括数千个个体，但在结束时只在小区中鉴定几百个家系， 而在正式的大田鉴定中只有更少的家系。 因此，系谱育种的第一个重要问题是如何有效地实施这一选择过程， 特别是对受显性效应或环境强烈影响的性状 (Simmonds 1979)。第二个问题是当系谱育种的目标是为了获得某特定性状的超亲变异时，要确定性状表现相当的栽培品种是具有类似的还是互补的遗传结构，即是否对于某些基因座位，正向（增效）等位基因仅存在于一个品种中，负向（减效）等位基因则存在于另一品种中，而对另一些基因座位则相反，也即存在同效等位基因的分散分布。只有当存在互补的遗传结构时才可能获得超亲的重组体。Lande(1992) 认为在系谱选择计划的早期世代，标记辅助的选择可能是非常有效的。

10.2.1 单标记选择

与回交一样，近交期间标记选择的有效性由 M_1 和 Q_1 之间连锁的紧密程度决定。Dudley (1993) 推算出从双杂合 F_1 开始，在不加选择、在 F_2 中选择 M_1M_1 纯合体、选择至少带有一个 M_1 等位基因的个体三种情形下纯合体 $M_1M_1Q_1Q_1$ 所占的比例（表10.2）。当 r =0.01，在 F_2 选择标记纯合体时，99% 的 F_6 品系将是理想QTL等位基因的纯合体。如果有4个标记座位处于分离状态，每一个与一个QTL相连锁且 r =0.01，则96% 的品系将在4个座位上均是有利 QTL等位基因的纯合体。但是，如果 r =0.1，90% 的品系为单个 QTL的纯合体，只有65.6% 的品系为四个 QTL 的纯合体。因此，基因座位数和 r 两者都影响着获得理想基因型的概率。

10.2.2 多标记选择

利用单一标记鉴别和操纵理想目标基因，要求它们之间存在紧密连锁（标记和目标基

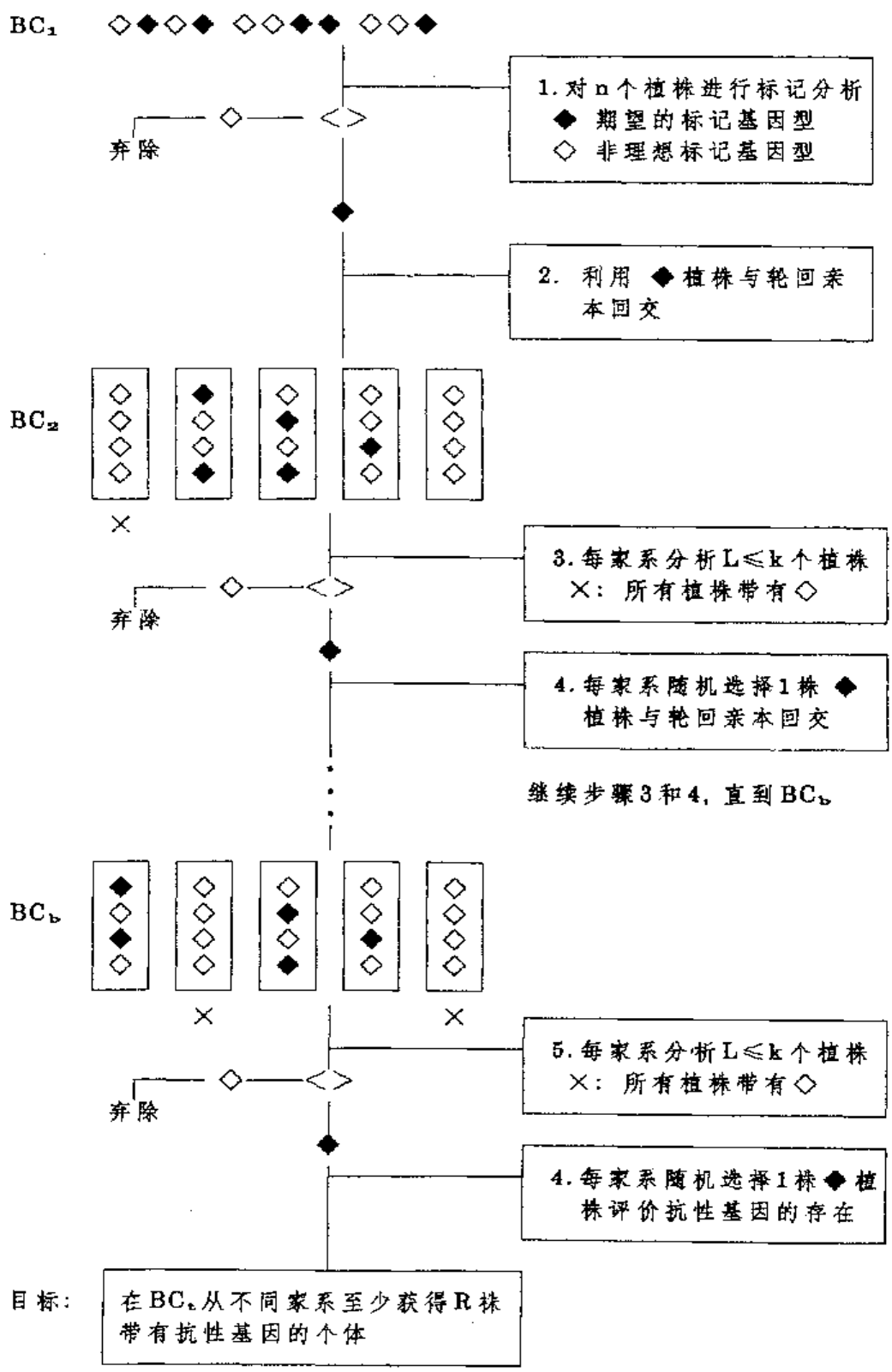


图10.2　回交育种计划中标记辅助的单基因抗性选择

因座位之间的距离小于5cM)。当连锁松弛时，由于标记和目标基因之间的重组率较高，利用标记进行操纵的效果就会被抵销。但当目标座位由相邻标记所包括时，通过同时选择两个或更多相邻标记，就会大大提高借助于次紧密连锁标记所进行的基因操纵的效率。只有当两个相邻标记之间的区域发生双交换时，才无法检测到目标座位有利等位基因的丢失。为了有99%的把握找到目标基因,Tanksley认为整个基因组标记的间隔不应大于20cM。这

表10.2　从相引态F_1获得基因型为$M_1M_1Q_1Q_1$的F_2家系的比例

重组率	F_2选择方式		
	不加选择	选择M_1-	选择M_1M_1
0.50	0.2500	0.3333	0.50
0.40	0.2778	0.3704	0.60
0.30	0.3125	0.4167	0.70
0.20	0.3571	0.4762	0.80
0.10	0.4165	0.5554	0.90
0.01	0.4894	0.6525	0.99
0.00	0.4990	0.6654	1.00

样的饱和度和分布实质上在玉米和番茄中早已经达到(Helentjaris等1985,1986b; Tanksley和Hewitt 1988)。但这种间隔的同工酶标记在任何栽培植物都是难以达到的。

10.2.3　多个性状选择

迄今为止，大多数研究只是集中于对单个性状进行标记辅助选择。当标记效应对一个以上的性状显著时，沿着理想的方向对影响各个性状的等位基因进行选择可以同步改良每一性状。另外，同步改良一个以上性状的选择指数方法也适用于分子标记辅助的选择。这些方法各有其优缺点。利用对两个性状均具有正效应的那些座位可以使两个性状都得到改良。但是，它失去了单个性状的潜在获得量，并且可能限制净经济值(Net worth)的进展。为了评价在多个性状选择上的潜力，要考虑到对于一个性状，特定座位的测定值是性状基因型均值对基因型的回归系数。对于每一性状，每一标记等位基因可以给定一个基因型值，这取决于该标记座位的回归。这些值然后用每一性状的经济值进行加权。如果$r=0$，每一性状的遗传率就是1。因此，回归系数矩阵与经济值的乘积就提供了一组权重，这些权重分配给特定植株或子代每一座位的分子标记基因型。这些权重之和可能就是特定植株的指数值或净优值(Net merit)。另一可供选择的指数值是根据表型值确定的，可按每一植株或家系计算出来。该值可处理为一个性状，并可鉴别该指数的QTL与标记的关联。

10.2.4　寡基因抗性的选择

在育种工作中，与抗性基因紧密连锁的标记可用于单基因抗性的间接选择。分子标记在抗性筛选和鉴别方面还具有以下特别的应用。

1. 替代不可靠的自然鉴定

植物的抗逆性由于在田间难以均一地进行鉴定，使选择进程受阻。而且，人工环境的设置既困难又昂贵。在这种情况下，标记分析比抗性试验更可靠和简便，并以其可接受的成本，为大规模筛选育种材料提供了可能。例如 Rick和Fobes(1974)提出番茄的抗线虫基因 *Mi*，可借助其与同工酶基因座位 *Aps-1* 的紧密连锁来进行选择。许多种子公司已利用该标记将 *Mi* 抗线虫基因转育到育种材料中。

2. 对外来或作为检疫对象的病虫草害实施选择

有些容易流行的病虫草害不易在受试的环境下进行自然鉴定，从安全起见，常禁止人工模拟其发生过程。但是，在许多情况中育种者希望对抗性进行选择。

在上述情况下，借助标记选择是筛选抗性的唯一可能途径。根据这一点，国际玉米和小麦改良中心 (CIMMYT) 最近正致力于检测与玉米霜霉病抗性和穗条纹病抗性连锁的分子标记。因为在一些非洲国家的自然条件下，穗条纹病的发生无规律，故借助标记的选择可以保证在非洲育种圃或其它地区非正季圃中生长的材料的抗性得到同步改良。

3. 避免环境效应对抗性选择的干扰

许多抗性基因的表达受环境条件的强烈影响。如谷类作物的条锈病和茎锈病，其病症会受到温度的影响。借助标记的选择克服了抗豌豆母体花叶病毒 (PEMV) 育种中碰到的类似难题。在高温下，感病植株的病症消失，而在低温下，抗性植株上出现类似致病症状。Weeden 和 Provvidenti(1988) 发现抗 PEMV基因与一个同工酶基因座位高度连锁(约4cM)，并认为借助标记的选择将比直接接种筛选更加可靠。

4. 抗性的早期选择

一些抗性表型直到最后发育期才表达，而分子标记通常可在很早的生育期进行。故待进行直接抗性测定时，无需保持大量的植株。对于世代周期长和栽培成本高或形态高大的物种生育后期表达的抗性，在幼苗时借助于标记基因进行早期选择是非常有益的。例如，加州山松泡锈病的抗性在3～5年龄的植株上才能直接选择，而分子标记辅助的选择在早期就能进行。

5. 抗性基因的累积

抗性基因的累积已作为获得流行或非流行性病原菌小种持久抗性的一种策略，它是指把抗同一种病原菌的不同基因集中到一个品种（系）中。每个抗性基因要么为有效的（使病原菌丧失致病性），要么为无效的(其抗性已为病原菌所克服）。实质上，对于累积基因产生更为稳定的抗性有两种观点，它们取决于基因有效性还是抗性的丧失。在抗性基因的有效性（即基因控制着所有现存小种的抗性）方面，Schafer和Roelfs(1985) 的计算表明一个病原菌克服由4～6个基因累积的抗性的概率非常之低。在无效抗性基因产生的残余抗性方面，Nelson (1979) 认为由众多抗性基因的部分效应累积所产生的抗性基因型几乎不受到病原菌的选择压力，并且因此其抗性将会更加持久。尽管这种概念仍未被广泛接受，但有实验证实了某些寄主与寄生物体系中存在残余抗性效应。

类似地，可能将对多种虫害或同一虫害的多种生物型的抗性基因通过分子标记辅助的

选择转移到同一栽培品种之中，使抗性基因得以累加，抗性得以增强或持久。

在番茄中，现已建立了多个基因与 RFLP 标记的紧密连锁关系。这些基因包括抗烟草花叶病毒 (Young 等1988)、枯萎病、细菌性斑点病和根节线虫的基因，以及控制番茄株型和果实成熟特性的主基因 *sp* 和 *u* (Paterson 等1988)。在玉米中建立了 RFLP标记与抗玉米矮缩花叶病毒基因的连锁关系。莴苣中也已确定了 RFLP 标记与抗霜毒病基因的连锁关系 (Landry等1987) 。目前在栽培物种中分子标记与主基因之间的紧密连锁关系已有不少报道，从而为标记辅助的杂交育种和近交选择提供了基础。

10.3 图式基因型和全基因组选择

在有性繁殖的生物中，染色体是一系列世代所衍生而来的 DNA 嵌合体。在减数分裂中，一对特定的同源染色体之间可能发生零次、一次或多次的交换事件。当交换发生以后，要完整地描述一个基因组，就必须了解由于重组所造成的等位基因组成的变化，以及重组事件所发生的区域。尽管有时传统的核型分析也能提供这类信息(Tease 1978)，但一般来说从简单的显微核型分析不能得出特定染色体或染色体片段来源的信息。相反，在传递遗传学领域，经常要描述各个个体在一个或多个基因座位的基因型。这种描述一般是采用字母和数字来表示的，以提供特定基因座位的衍生过程和等位基因组成。但是，在过去所进行的大多数传递遗传学实验中，一次只能跟踪少数几个基因座位。

高密度的分子标记图谱可用于同时确定各个体在许多甚至数百个基因座位上的分子标记的基因型。因此，分子标记分析可用于推测特定个体在整个基因组上任一区域最为可能的遗传组成。但是随着所分析的基因座位数的增加，要描述各个基因座位的数字式基因型就变得极为困难。Young 和Tanksley (1989a) 发展了一种以 RFLP分析为基础的图式基因型分析的方法。它以图式形式反映RFLP基因型，较之数字式基因型有许多优点。这种图式基因型类似于在一次图象分析中所得的描述整个基因组的染色体核型，但所不同的是图式基因型是根据 RFLP 数据推测的，并且因此能显示基因组任意一点上的基因组组成及其亲本来源。在实践上，根据 RFLP分析所确定的图式基因型可用于在复杂人类遗传病的预测中简化多基因性状的分析 (Lander和 Botstein 1986) 以及在植物和动物的多基因性状育种中实施全基因组选择 (Whole genome selection) (Osborn等1987; Tanksley和Hewitt 1988)。

10.3.1 推测图式基因型的基本前提

基本前提条件包括：已建立完整的分子遗传图谱并已知分子标记的顺反构型。由近交系衍生的群体，从培育程序就可推测其顺反构型。有时情形比较复杂，必需从三个世代获得完整的 RFLP 资料，才能确定第三代个体的图式基因型。例如在人类方面，为了获得系谱中子代的 RFLP 图式基因型，必须分析祖父母、父母的 RFLP 数据。如果没有顺反构型的信息，基因组某些区域的 RFLP 资料就可能具有一种以上的图式基因型，而且它们具有同等可能的正确性。事实上，由 F_1 杂种自交产生的 F_2 群体可能含有一部分固有的模棱两可图式基因型。

10.3.2 推测图式基因型的基本假定

在推测图式基因型时，两个标记区间的基因型从限定该区间的标记的基因型推测而来。当从 RFLP 端点的基因型推测某区间的图式基因型时，通常具有不同的构型可供选择，这些构型均可以满足可供利用的 RFLP 资料。制作图式基因型的原则是利用最为可能的构型。因此，在制作图式基因型时所利用的是所需交换事件最少的简单构型，而不利用需要一个或多个多重交换事件的其它构型。 在实践上，如果两个相邻的基因座位具有相同的基因型，则它们之间染色体片段的基因型就推测为这两个标记的基因型。当两个相邻的基因座位 RFLP 基因型不同时，就认为在这两个标记之间的某一区域发生了染色体片段的交换。

由于非重组体区间的基因型是从其 RFLP 标记端点的基因型推测的，因此在某一特定区间如果存在双交换或偶数交换，就证明这种推断是错误的，并且双交换的概率增加量为两个相邻标记间交换率的平方。因此，对于任意区间，所推测的基因型为正确的概率为$1-r^2$，其中r是两个相邻标记间的交换率。对于整个基因组，所有区间推测正确的概率为

$$Pr = \prod_{n=1}^{\text{总区间数}} (1-r_n^2)$$

上式只考虑了双交换，且忽略交换干扰。 若考虑一个基因组总长为1000cM、RFLP均匀分布的生物，从 RFLP 基因型所制作的图式基因型完全正确(即图式基因型推测错误的区间数为零)的概率如图10.3所示。 图10.3还显示了图式基因型所能正确描述的基因组所占的期望比例。计算这一值时先确定特定在 RFLP 标记的特定间距条件下描述错误的区间数为0、1、2、... 的概率。 这些概率与间距的大小一起用于确定可以正确推测的基因组的期望长度，然后除以基因组的总长度(1000cM)，就得到该图式基因型所能正确描述的基因组的期望比例。从图10.3可以看出，利用相互间隔距离为10cM 的 RFLP 标记，所推测的图式基因型在所有区间完全正确(推测错误的区间数为0)的概率只有33%(曲线B)。但这一图式基因型在描述基因组构成方面对于99%以上的基因组都是准确的。 即使 RFLP标记间的距离增加到30cM，所推测的图式基因型对于约95%的基因组也是准确的(曲线A)。

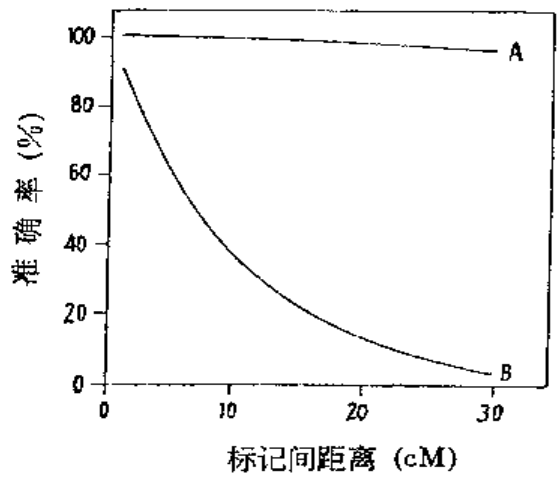


图10.3 图式基因型准确性的理论估计

10.3.3 图式基因型实例

Young 和 Tanksley(1989a)给出了栽培番茄和野生番茄之间 F_2 个体 RFLP基因型的数字数据(图10.4a)，相应的图式基因型列于图10.4b。在 F_2 情形下，每一基因座位有三种可能的基因型。图10.4a为 RFLP 基因型的数字数据(共分析了70个 RFLP 基因座位)，按基因组顺序排列，1表示栽培番茄的纯合体，2表示杂合体，3代表野生番茄的纯合体。 如前所述，F_2 个体可能有本来就不确定的基因组片段。图10.4b为从数字基因型推测的图式基因型，空白区域代表来自栽培番茄的染色体片段，阴影区域代表含有交换事件的片段；黑色区域代表来自野生番茄的片段。每一对染色体都显示了两个同源染色体的图式基因型。同时给出了具有同等可能性的两个图式基因型异构体，其差异所在区域用箭头和实线标出。当多个杂合基因座位被一个或多个纯合基因座位分开时，这种两可性就会发生。这时就有两个具有同等可能性的图式基因型，表现出相邻杂合区域顺反构型的差异。但根据泊松分布所作的计算表明，对于含有10个染色体每个长100cM 的基因组，只有6% 的基因组区域是模棱两可的。 因此，利用 F_2 群体进行图式基因型分析一般不会受到顺反构型两可性的严重影响。

10.3.4 利用图式基因型进行全基因组选择

在育种中，由于分离后代植株的染色体是亲本染色体片段的嵌合体，故选择不但要考

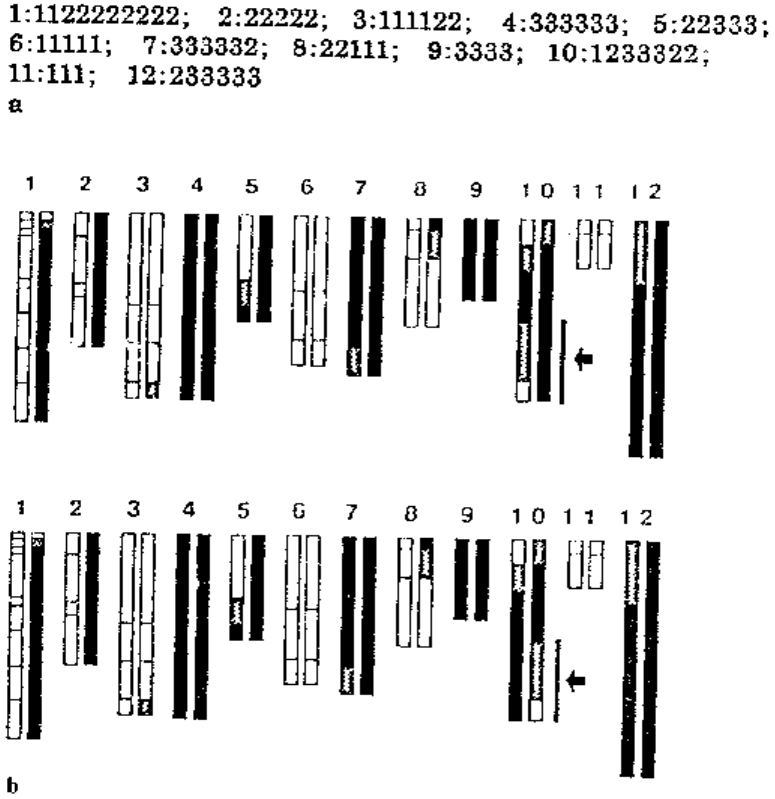


图10.4 一个 F_2 个体的图式基因型

虑目标基因，而且也要考虑植株的整个基因组。如果弄清了某一特定基因座位的 RFLP 基因型，我们就可推测基因组中该特定座位上的等位基因来源于哪一个亲本。一旦知道了整个基因组许多连锁座位上的 RFLP 基因型，我们就可根据亲本染色体来推测每一个体染色体的确切组成，即根据基因组中相关连锁基因座位的信息推测出连续的基因型。图式基因型应用于选择的主要优点在于它使得有可能进行全基因组选择。换言之，它使得有可能理解、选择和操纵特定个体的整个基因型。

1. 回交选择

举例来说，如果已知番茄对一些病虫的抗性是由位于染色体上已知区域（图10.5中箭头所指之处）的三个基因座位控制的。图式基因型就可以用于对这些基因进行选择。在所有三个座位上，带有B亲本等位基因（黑色表示）的个体表现抗性，但基因组其它区域的B亲本等位基因控制着非理想性状。育种的目标就是要获得在三个抗性基因座位上等位基因为B亲本类型，而其他区域为A亲本类型（空白表示）的个体。利用带有B亲本三个抗性等位基因的

第1植株

1: 1111222222; 2: 22221; 3: 222222; 4: 111122; 5: 22222; 6: 21122; 7: 111111; 8: 11112; 9: 2222; 10: 2222222; 11: 112; 12: 12222

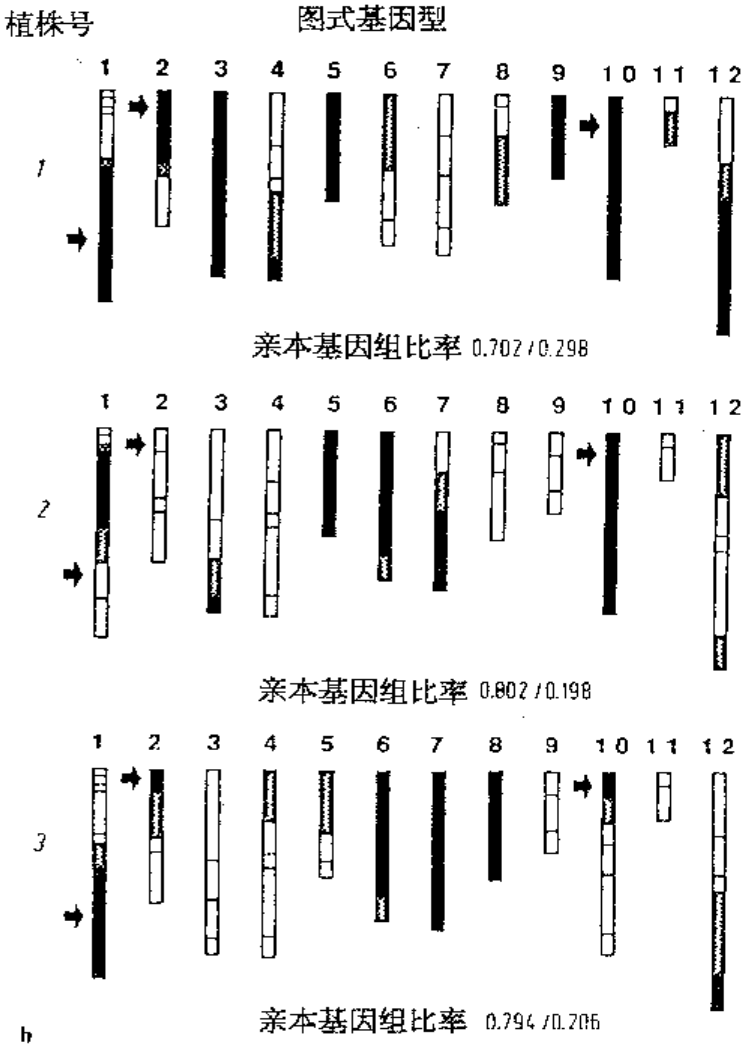
第2植株

1: 1122222111; 2: 11111; 3: 111222; 4: 111111; 5: 22222; 6: 22211; 7: 112222; 8: 11111; 9: 1111; 10. 2222222; 11: 111; 12: 21112

第3植株

1: 1111122222; 2: 22111; 3: 111111; 4: 211111; 5: 21111; 6: 22211; 7: 222222; 8: 22222; 9: 1111; 10: 2222111; 11: 111; 12: 11112

a



b

图10.5 番茄回交群体三个个体的图式基因型
略去了全部由轮回亲本DNA 组成的另一同源染色体

个体与轮回亲本A反复回交，直到B亲本的非理想DNA 被去除时为止。 这样就可以最终获得理想的抗病个体。

比较图10.5中的三个个体，可以很快找到最适合于回交的个体。 在三个个体中，1号和3号植株在三个基因座位上都保留了亲本B的抗性等位基因。育种目标要求同时去除亲本B的非理想等位基因。比较1号和3号植株发现，3号植株在所有三个目标基因附近都发生了重组事件，因而去除了来自亲本B的大量侧连 DNA。相反,1号植株在此三个基因座位附近则保留了大量来自亲本B的DNA 片段。 而且亲本基因组的比率(个体中分别来自亲本A和B的基因组之比)表明3号植株的总 DNA中来自亲本B的 DNA只占20.6%，而在1号植株中则占29.8%。 因此在下一轮回交中以3号植株作亲本，获得理想基因组个体所需的回交世代数较少。

2. 自交后代的选择

如果两个不同遗传材料各有其优缺点，它们在许多性状上是互补的。 理想的育种计划是通过杂交，在其自交后代中选择具有双方优点的重组体。通过分子标记与这些理想性状间的连锁关系，可以根据各个体的图式基因型来提高选择的效率。与回交选择所不同的是，除了选择在这些目标基因座位上具有理想的结合形式外，基因组的其它区域可以让其随机分离和重组。所要综合的目标基因座位越多，选择的效率就越高。 这对于抗病性、抗虫性以及各种逆境抗性等多性状选择具有更为重要的意义。

Young和 Tanksley (1989a)开发了一个适合于回交和F_2群体、在 Apple-Macintosh机上使用的计算机软件，可用于将群体中各个体的数字式基因型转换为图式基因型并确定亲本的基因组比率。此外，该软件还可根据用户给定的参数对分离群体中的特定基因组类型进行选择试验。 徐吉臣等 (1995)利用这个软件分析了水稻窄叶青8号和京系17之间的DH 群体，并确定了几个基因组区域与农艺性状的关联。

10.4 选择效果及其评价

在某个分离群体中准确鉴别的与有利的 QTL等位基因相关联的标记等位基因是否可以用于在不同分离群体中对该 QTL有利基因的选择，将取决于：(1)待选群体中标记和QTL分离的存在；(2)两个群体中标记和 QTL等位基因之间的连锁相(相斥或相引)是否相同；(3)当存在上位性时其它 QTL的基因型。如果标记和QTL都是杂合的、在两个群体中连锁相相同，且上位性不重要， 则在一个群体中鉴别的关联就可以用于另一个群体的选择。Bubeck 等(1993)发现在三个玉米群体中只有一个染色体区域是一致性地与叶斑病抗性相关联。但是，群体间的差异与环境所造成的差异相混淆。Beavis等(1991)惊奇地发现少数株高的 QTL在四个玉米群体间是共有的。但是，他们所采用的接受标记－QTL关联为真实的概率水平，以致当 QTL 存在时不鉴别QTL 的概率为0.5。因此，群体间一些共有的QTL可能被丢失了。

10.4.1 玉米群体数量性状的选择响应

Stuber等 (1980) 报道了玉米长期选择计划中，同工酶等位基因频率的变化与籽粒产量之间的显著相关。由这些结果可以假设，对同工酶标记座位的等位基因频率进行操纵可导致相关性状的显著响应。在异花授粉的玉米群体中为检验这一假设所进行的研究表明，

依赖于7个同工酶座位基因频率的操纵所作的选择显著地提高了产量，与籽粒产量高度相关的穗数也随之增加。在一项类似的研究中，以优良自交系的复交组合所产生的群体为材料，同工酶频率操纵所产生的响应与对同群体进行表型选择所得响应几乎相等（Frei等1986b）。在这两项研究中，由于目标群体在研究前经过了几个世代的随机交配，因此，标记座位与 QTL 之间连锁不平衡的程度可能相对较低。这无疑降低了利用标记座位操纵相关性状的有效性。尽管这些研究曾被认为是中等成功的，但其结果将足以激励我们利用大量标记和不同类型的群体对数量性状作进一步的研究和操纵。

Stuber 和Edwards (1986)报道了利用标记选择程序操纵玉米数量性状的深入研究。COTX 和 CMT 两个 F_2 群体的结果为这一研究提供了基础。在早期研究中选择只是根据 F_2的基因型，对其异花授粉植株的后代进行选择响应的评价。在另一项研究中，以 COTX 和 CMT F_2 单株的另一样本为材料，选择得到自交子代。根据每个 F_2单株的表现型进行表型（混合）选择，便于与标记辅助的选择响应进行比较。在单个 F_2植株研究的次年，在北卡罗来纳三个地点进行所有选择响应的评价。

在每项标记辅助的选择研究中，测定了每个标记、每株F_2全部性状的育种值。育种值定义为标记座位两种纯合体类型数量性状平均值之差的一半。有利标记等位基因处于纯合状态的植株赋以正值，不利等位基因纯合的植株赋以负值，杂合体取值为0。将每一植株每一性状15个同工酶标记座位的育种值相加，得到每个单株/性状组合的复合育种值。利用复合育种值对籽粒产量、穗数和穗位高度各性状进行选择，组成正向和负向的歧化选择群体。

对于 CMT F_2群体中标记辅助的基因型选择，增加产量（正向）的个体类群，比减少产量（负向）的个体类群平均值高40%。而且，正向个体类群的平均产量比未选择的对照（同一 F_2的异花授粉群体）的平均值高20%。增加产量的表型选择响应几乎等于基因型选择响应。 COTX F_2群体中基因型选择的响应类似于CMT F_2，但是产量的变化没有如此巨大。这一点是可以预期的，因为标记座位解释的产量表型变异，在CMT组合中大大高于 COTX 组合。穗位高度和穗数与籽粒产量存在高度相关，并且这两个性状的选择响应也反映了这种高度相关。自交子代的选择响应结果与之类似。

由于已确定了每个选择单株的标记基因型，可以计算每一选择群体标记等位基因的频率。对于 CMT 群体，标记辅助的基因型选择，向提高产量方向选择的群体中15个同工酶座位有利等位基因的频率平均比降低产量选择的群体高0.38。歧化选择所导致的单个标记座位等位基因频率的差异为0.02～0.73。正如所预期的那样，解释最大比例产量表型变异的那些标记座位，其等位基因频率的差异也最大。

尽管对于经表型（混合）选择所衍生的群体而言，产量的基因型选择和表型选择获得了类似的选择响应；但对于以标记为基础的选择，歧化群体间基因频率的差异平均为0.13，约为0.38的1/3。这意味着，在标记选择程序中未经受选择压力的基因组非标记区域的基因座位对表型选择响应产生了影响。 COTX 群体中的结果与 CMT 类似，但频率差值小于COTX群体。

这些研究表明，根据可能代表不到 30～40% 基因组的 15个同工酶标记座位的标记选择结果与包括整个基因组的表型选择一样有效。而且，如果整个基因组为均匀分布的标记（间隔为 10～20 cM）所覆盖，则上述结果就可能意味着标记选择的相对效率会显著增加。由这些结果还可以推测对于产量等复杂性状的操纵，标记技术可在品种间有效地进行多个标记座位理想因子的转移。这种操纵和转移的效果和成功与否将随遗传背景而变化。

10.4.2 长期选择结果的评价

分子标记也可以用于对育种计划的选择结果进行评价，从而在分子水平上了解表型选择的效果。众所周知的一项选择计划是美国Illinois 大学在1896年开始对玉米油分和蛋白质含量等进行的选择。Sughroue和Rocheford (1994)对从不同方向选择所得品系进行了RFLP 分析，发现品系内和不同品系间均存在相当大的多态性。品系间 RFLP 复等位基因/突变体较为普遍，其中59% 的 RFLP 座位表现为多重突变。Goldman等(1993)对不同类型品系间的100个 F_3 家系采用90个基因组和 cDNA 克隆进行检测，发现了100个多态性座位。单因素方差分析表明，分布在10个染色体臂上的22个座位与蛋白质含量存在显著关联，在染色体臂 3L、5S和7L 上检测到3簇或3簇以上与蛋白质含量存在显著关联的座位，表明在这些座位上存在较大效应的 QTL。由不同染色体上的6个显著 RFLP座位所组成的多元线性回归模型可解释蛋白质含量变异的64%。因此，几个具有较大效应的染色体区域可能控制着这些品系间蛋白质含量的大部分变化。

10.5 杂种优势分析

当两个优良栽培品种杂交时，F_1 常常在总体经济价值上超过较好的亲本品系，偶尔也会在特定数量性状上超过较好的亲本。这种杂种优势现象目前主要采用经济性状正向（增效）等位基因的直接显性效应和亲本品系遗传结构的互补来解释。在 F_1 中，当一个品系的正向（增效）显性等位基因遮盖了另一亲本的负向（减效）隐性等位基因时就会产生优势的杂种，反之亦然。单个基因座位的超显性（杂合体优势）也会对杂种优势起作用。

10.5.1 杂种优势的分析

目前杂优组合的改良主要是将大量待测品系与亲本进行杂交，希望能够偶尔碰上一个比现有推广杂种优良的组合。在这一计划中所用的近交系过去都进行过一般配合力的筛选。但是，由于不同杂交种的表现呈现正态分布，故这类程序所得回报将很快降低到最小。也就是说，在检测了几百个组合之后，要找到一个略为超过目前最好组合的杂种，就得额外地多检测数百个组合。已倡导了一些改良杂种优势的其它途径（如交互轮回选择及其各种变形的轮回选择方法），但在实践上似乎并未得到应用。

标记－QTL 连锁信息为改良现有组合的杂种优势提供了另一途径。这一途径为下列研究提供了手段：(1)从潜在种质资源中系统鉴别与现有杂交组合的亲本之一表现互补显性的其它染色体区域。(2) 将这些染色体区域渐渗导入相应的亲本品系。在这一程序中，首先对现有组合的 F_2 进行标记－QTL 连锁分析。目的是要鉴别有利效应表现为显性甚至超显性的标记连锁的染色体区域。这些区域可通过标记分析中对杂合标记基因型数量性状值的考查来检测。这些区域对推广组合 F_1 杂种优势效应起作用。近交种质品系与亲本品系之一杂交，并进行新的标记－QTL 连锁分析，以检测该种质品系中与同源的亲本染色体区域表现正向（增效）显性的其它标记连锁的染色体区域。一旦鉴别出来，采用适当的标记来监测渐渗过程就可将这类染色体区域从种质品系渐渗到另一个推广的亲本品系中。重复进行上述过程，即可有计划地增加两个亲本品系间互补显性的程度，产生杂种优势更强的 F_1。

10.5.2 杂种优势的预测

在玉米中，采用传统的育种方法来培育可产生杂种优势的自交系既费成本又花时间。为了筛选配合力高的自交系，大部分时间都花在对这些自交系所产生的单交组合的田间评价上。因此，在配制杂交组合或对成百上千的杂交组合进行试验之前预测杂种优势的可靠方法，对玉米和其它作物的杂种生产是极为有益的。几个研究者已估计过同工酶等位基因多样性与玉米自交系单交组合杂种产量之间的相关性(Hunter和Kannenberg 1971; Heidrich -Sobrinho和 Cordeiro 1975; Gonella和Peterson1978; Hadjinov等1982)。这些研究采用11个或11个以下的同工酶基因座位，15或15个以下的自交系，所估计的同工酶多样性与特殊配合力之间的相关是较低的或不显著的。即使是在采用了从37个优良自交系产生的100个玉米杂种、用30个同工酶评价相关性的研究中，Smith和Smith(1989)将F_1产量对同工酶等位基因多样性作图所得 R^2也只有0.36。在对75个 F_1水稻杂种进行6个同工酶基因座位的研究中，也未发现杂种籽粒产量与同工酶多样性之间的关联 (Peng 等 1988) 。

Frei等 (1986a) 报道了对114个玉米单交杂种研究。他们比较了利用同工酶标记多样性和利用系谱多样性预测杂种产量时的各自优点。根据21个同工酶座位的基因型，将成对自交系分成相似和不相似两大类。然后再根据成对自交系间系谱背景是否相同将每一同工酶多样性类群进一步细分为系谱相似和不相似两个亚类。不相似同工酶类群的杂种平均产量显著高于（高10%）具有相似同工酶的类群。具有不相似系谱的杂种产量则比相似系谱的杂种产量高37%。研究者认为尽管在单交组合中同工酶标记的不相似性与较高产量存在显著关联，但这些标记的预测价值主要限于具有相同系谱的自交系。Lamkey等 (1987) 研究了 Iowa Stiff Stalk Synthetic (SSS) 群体经一粒传产生的247个自交系中24个高产和21个低产自交系间的单交杂种，比较了这些高低产自交系间高×高、高×低、低×低三类杂种的产量。结果表明，利用11个同工酶标记座位的等位基因差异不能预测杂种的表现。

根据以上研究，同工酶在玉米和水稻杂种表现预测中的价值是有限的。令人失望的结论可归结于几个因素。例如，在大多数研究中只是分析了数量较少的同工酶基因座位，只能有效地覆盖基因组的小部分。结果，只有有限比例的遗传因子的抽样是有效的。更为重要的是，大多数标记座位的等位基因可能并不直接影响数量性状的表型表达，而是用于鉴别相邻的染色体片段。当可用于杂种表现预测的标记有限时，必须明确与特定标记等位基因连锁的 QTL 的效应。标记座位的等位基因差异并不能保证相邻QTL 也存在等位基因差异。而且，如果被测杂种是从随机交配群体产生的自交系衍生而来的，或者这些自交系包括可公用的自交系中的某一亚套，则可预期标记等位基因与 QTL等位基因之间的关联实质上是随机的，即接近于连锁平衡。为了将以标记为基础的方法成功地应用于籽粒产量等复杂遗传性状的预测或选择，基因组必须被均匀分布的标记所充分饱和，或者必须存在高度的连锁非平衡。

在Smith和Smith (1989) 的研究中，利用37个优良自交系所衍生的100多个杂种，测定了产量与 RFLP 基因型多样性之间的关联。根据230个标记，将 F_1籽粒产量对 RFLP多样性作图，所得 R^2 值为0.87。因此，利用230个RFLP座位与利用31个同工酶座位(R^2为0.36)两种结果形成鲜明对比。即使是利用230个 RFLP 标记，亲本间具有最大距离 (0.70～0.80) 的杂种，其籽粒产量也变化在8.15～10.66吨/公顷。

10.5.3 与杂种优势有关的QTL

Stuber 等(1992)利用分子标记鉴别了与玉米杂种优势有关的遗传因子。他们发现，与籽粒产量有关的 QTL，其杂合体通常比对应的纯合体具有更高的产量，表明除超显性外，这些 QTL 在杂种优势形成中也起着重要作用。De Vicente 和 Tanksley (1993)则利用野生番茄和栽培番茄的 F_2群体分析了与种间杂种超亲分离有关的 QTL，发现 76 个显著的 QTL 中，36% 的 QTL 其等位基因的效应与根据亲本表型预测的效应相反，这些 QTL与F_2群体中超亲个体的表型有直接的关系，只有干重累积量是个例外。将两个与干重累积量有关的超显性 QTL回交转移到栽培番茄中，获得了超亲个体，并且超亲个体的出现与这两个 QTL 有关，表明超显性是超亲分离的第二个原因。以上结果表明，分子标记可以用于鉴别与超亲分离有关的 QTL，并可选择性地将其转移到栽培品种中。

10.6 无性繁殖品种的改良

无性繁殖品种一般是以经济价值较高的单个个体为基础的，这些个体来自于正常异花授粉群体中独特的历史性遗传事件。无性繁殖品系的改良，一般要将特定无性系中的某一个体自交，或者将不同无性系的两个个体杂交，以求获得优良的 F_1重组体。由于亲本单株通常是高度杂合的，优良的杂合子代可采用无性系的形式进行繁殖，因此没有必要自交到 F_2来寻找重组体。正如系谱育种那样，可以从一个非常大的F_1重组群体开始，通过连续选择逐渐减少品系数并增加测验的准确性，直到只保留少量的品系来进行大规模的鉴定，而且象系谱育种那样，主要问题是要在早期阶段鉴别出具有形成优良无性繁殖品种潜力的重组型个体。

标记－QTL 分析可以提供与优势经济价值有关的标记连锁染色体区域的信息，因而有助于上述问题的解决。正如象系谱育种那样，根据显著受环境影响的性状的标记组成可以选择用于精细鉴定的候选材料。但应该指出的是，一旦获得有关标记－QTL 连锁关系的信息，就可以培育很大的 F_1幼苗群体，并根据标记组成对其进行选择或者再根据标记组成进行几轮选择，以提高保留下来用于更精细鉴定的个体中优良基因的频率。

10.7 生物起源与进化的研究

植物遗传学的发展主要是围绕重要栽培植物的研究而发展起来的，由于大多数作物不能进行种间相互杂交，其遗传学的发展或多或少具有独立性、且不相互交叉。它们各自有自己的研究人员，有时还有专门的杂志。

尽管许多作物看上去是独立的，但它们还是具有分类学上的亲缘关系。例如，甘蓝、芜菁和油菜都属于芸薹属，玉米和高粱都是禾谷类。番茄、马铃薯和胡椒均是茄科植物，它们的染色体数是相同的 (X=12)，但相互之间不能杂交，迄今有关它们基因组之间染色体的同源性知之甚少。如果染色体的成分和基因顺序在亲缘种中同源性很大，这就有可能（通过体细胞杂交）置换单个染色体或染色体片段，从而把常规的种内杂交所不能得到的那些作物特征结合起来。

用cDNA 克隆作探针构建遗传图谱赋于 RFLP 连锁图功能的意义，因为图谱上的每个

标记都是表达基因的座位。若能最终阐明每种cDNA 编码的蛋白质功能，这样的 RFLP 图谱就成为基因组的基因分布图了，当然这是一个长远的基因组研究目标。但现阶段可利用基因的编码序列在进化上保守的特点，用同一套 cDNA 克隆（或基因组的随机单拷贝序列库）作探针构建同属或同科的近缘物种的 RFLP 图谱，进行基因组结构和功能的比较，实验表明以燕麦的cDNA 克隆可对禾本科的小麦、大麦、水稻、甘蔗等各种作物进行 RFLP 分析。

现以茄科为例说明这方面的研究进展(朱立煌 1992)。茄科植物包括象番茄、马铃薯、辣椒、茄子和烟草等重要的蔬菜和经济作物，它们都有12对染色体，但彼此间不能杂交。Tanksley实验室首先建立了番茄的RFLP 图谱，然后用已定位的 RFLP 标记在马铃薯和辣椒上构建图谱。对这三张图谱的比较引出了意想不到的结论，第一，在番茄和马铃薯的基因组中，几乎所有的标记都按一致的线性次序排列在同源染色体上，未曾发现在染色体间易位的迹象，只有在三条染色体上有不涉及着丝点的染色体倒位，这说明番茄和马铃薯在进化过程中基因组的功能结构未曾发生重大的变化 (Boniebale等1988)。第二，在番茄和辣椒中，基因组在进化上的保守格局被打破了，尽管两者仍然拥有同源的基因编码序列，但 RFLP 标记在染色体上的共线性排列因为大量染色体的重排而改变了。而且在辣椒的进化过程中有12%的cDNA 座位发生过加倍，从单拷贝扩增成多拷贝，并散布到不同的染色体上 (Tanksley等1988)。番茄和马铃薯之间非常相似的连锁群还表明，要在这两个种间相互替换染色体或染色体片段也许较为可行，并可增加保守区发生同源交换的可能性。相反，胡椒的 RFLP 图与番茄和马铃薯的 RFLP 图谱完全不同，故番茄与胡椒之间进行染色体替换的前景似乎较为暗淡。

再以马铃薯 patatin基因为例说明基因组比较与基因功能研究的关系。已知 patatin 是马铃薯块茎中主要的贮存蛋白，占可溶性蛋白的40%。在马铃薯中有10～15个patatin 基因，组成一个基因家族，所有的基因都有相同的编码序列，但根据与非翻译区是否含一段22bp的序列可将patatin 基因分成两型，patatin Ⅰ基因含22bp的序列，patatin Ⅱ则不含这一段。另外在Ⅰ型和Ⅱ型的 5′上游区还有其它的差异，这反映在表达的组织特异性上，Ⅰ型基因主要在块茎中表达；Ⅱ型基因在块茎中表达很弱，而在其它组织中正常表达。如前所述马铃薯和番茄的基因组具有相似的组织结构，但前者有块茎，后者无块茎，那么在这两种植物中patatin 基因座位的精细结构是否相同呢？为此 Ganal等用Ⅰ型和Ⅱ型基因序列作探针，通过 PFGE 对大片段DNA 进行电泳分析和物理定位(Ganal等1991)。结果发现在马铃薯第8对染色体长臂的端部有一个 patatin 基因的座位，含10～15个Ⅰ型和Ⅱ型的基因，其座位的物理长度在1.4×10^6bp范围内；在番茄第8对染色体相应的位置上也有一个patatin基因的座位，该座位不超过3×10^5bp，只含3个Ⅱ型的patatin基因。比较马铃薯和番茄的patatin基因座位的物理结构，可以清楚地看出这两种植物基因组的精细结构是与基因的功能一致的。另外在马铃薯和番茄的第3对染色体上都发现了一个只含patatin Ⅰ型基因启动子而不含编码序列的座位，这个座位是否有进化上的意义有待阐明。

10.8 其它应用

10.8.1 亲本控制和品种鉴别

分子标记在植物育种中最为直接的应用可能就是种子质量的监测和品系、品种的鉴定(Soller和Beckmann 1983)。特别是对于 F_1杂交种子，其生产和处理过程中发生的各种

错误和偶发事件可能使杂交种子因自交而导致较高比例的混杂。Arus (1983) 讨论了利用同工酶标记来检验 F_1 杂种的纯度问题。当亲本标记座位的等位基因处于纯合状态时，这种检验是特别有用的。

如果发现同一物种的大量品种间存在许多 RFLP，则某品种中所发现的片段的特定组合就为品系鉴别提供了独特的“片段指纹”。根据每一片段的多态性程度和复等位性，10～20个多态性谱带就足以满足大多数品种鉴别的需要。已证实人类的复等位高度可变区域为每一个体检验提供了独特的谱带模式。在植物中类似的探针可以提供鉴别单个栽培品系的独特“片段指纹”。的确，在玉米中发现与一个多基因家族杂交的单个探针可以将49个玉米自交系中的47个区别开来。转座子探针也可能提供独特的多条谱带模式。更不用说，插入 DNA 序列将在品系同一性鉴定中起重要作用，因为可以将插入序列加工使之具有独特而明显的人工特性。插入序列的存在可作为该品系起源的直接而明显的证据。

但是，在讨论利用同工酶标记进行品种鉴别问题时，Bailey (1983) 指出，由于其潜在的巨大分辨力，在将同工酶应用于育种专利时会产生更为广泛的问题。在同工酶差异上表现分离的一个现有品种可能被稍加改变而认为其具备某一同工酶表现型。或者一个受专利保护的品种，其电泳变异体被看作为稳定、受专利保护的新品种，尽管后者只是前者的一份“拷贝”。将 RFLP 用于品种鉴别时也存在类似的问题。

10.8.2 病原菌鉴别

在人类疾病防治中，RFLP 为特定细菌、真菌或病毒病原物的鉴别提供了简便的方法。同样，在植物中 RFLP 也可用于特定病害因子的诊断。例如，过去鉴别感染 *Fusarium* 物种的特定病原物基因型需要在田间对敏感型品系进行检验。现在，品系特异性 RFLP 模式为之提供了简便而直接的方法。在一项初步研究中，Manicom 通过与来自 *Fusarium oxysporum* f. sp. *dianthus* 的4个随机基因组探针杂交，对一系列 *Fusarium* 种进行 RFLP 筛选。发现有两个探针可以在物种 (forma specialis) 和分离物水平上揭示出差异。朱衡等(1994) 利用从日本稳定的稻瘟病菌株北1的基因组文库中筛选的重复顺序克隆，对17个稻瘟病菌株进行 Southern 分析，获得了可辨和重复的基因组特异的指纹图谱，并提供了8个稳定的稻瘟病菌株的致病性与其 DNA 指纹图谱之间的关系。在决定防治措施时，病原物类型的早期鉴定的意义是显而易见的。

10.8.3 利用外源种质的新途径

原始种和野生种存在大量遗传变异。外源种质中某一性状的变异程度常比栽培种大得多，因为栽培种常来自一个或少数几个祖先种，且人类已经对其进行了几个世纪的选择。

对某些性状来说，野生种质是基因的唯一来源。例如，在番茄中抗根节线虫的基因和抗烟草花叶病毒的基因都来自亲缘的种 *L. Peruvianum*。此外，在其它一些亲缘种内也找到了许多不同的抗病基因以及对逆境的耐性基因，但至今还未被利用。大多数其它作物的外源种质情况与之相似，如在小麦中，野生种质是抗许多不同病原菌抗性基因的来源，包括茎秆黄锈病、白粉病和叶锈病。小麦的野生近缘种中存在许多耐环境胁迫的基因，这些基因还有待于转移和利用。

植物育种家很早就认识到外源种质对品种改良的潜在价值。然而，他们通常不愿去利用这些有价值的资源，原因是野生种和栽培种难于杂交，且杂种通常部分不育或全部不育。

最近几年，组织和细胞培养技术已部分地解决了这一问题，但即使得到了可育杂种，也不能保证目标性状的基因可以成功地导入栽培种。把外源基因导入栽培品种的捷径是通过回交育种。如上所述，传统的回交不但花时间而且对转移多基因性状或遗传率低的性状无能为力。当转移外源基因时，连锁累赘的不利影响尤需注意。

由于采用分子标记能有效地选择含非必需供体DNA 最少的植株，故它们不仅能用于外源种质中目标性状基目的定位，而且可加速把含有利基因的少部分外源染色体 DNA转移至栽培品种。现在分子标记方法已成为把野生种的有利基因转入栽培植物的一种手段。

从长远角度看，利用分子标记检测来自外源种质的基因可能会成为该技术对植物育种最重要的贡献。通过把新基因导入栽培种的基因库，标记分析有可能改良一些过去由于缺少足够的遗传变异而很难改良的性状。

10.8.4 育种实践中的其它问题

分子标记的丰富程度及其在整个基因组上的分布为监测与育种有关的许多遗传过程提供了方法。这些过程包括：(1) 存在杂交障碍的种间杂种后代及其分离 (Lo Schiavo等 1983)；(2) 辐射后的部分花粉基因传递(Zamir 1983)；(3) 异花授粉作物基础群体中遗传变异的保持 (Orton 1983)；(4) 植株在不同条件下自交和异交的相对比例——这是育种家设计用于纯系保持、品系杂交、杂交种子生产、发展作物近缘野生种的抽样和保存方法等综合技术所需的重要信息 (Ritland 1983)。 例如在树木育种中，RFLP 已用于在人工杂交组合中监测自交和异交的相对比例以及来自外部花粉源的异交比例(Adams1983)。

10.8.5 植物群体生物学

同工酶在群体生物学许多领域的研究已取得很大进展，如群体遗传结构的分析和群体遗传变异的测定 (Brown和Weir1983)，自然选择的检测和评价 (Clegg 1983)，群体、亚种、种、属内系统发育和分类关系的确定，多倍体物种起源中同源多倍体化和异源多倍体化作用的研究 (Crawford1983) 。与同工酶相比，RFLP用于这些领域的最大优点是它们可检测非翻译序列和翻译序列的遗传变化。

10.8.6 无性系变异

分子标记可以发挥重要作用的一个新领域是分析无性系变异。无性系变异是指在离体繁殖材料中以较高频率出现的突变现象 (Karp和Bright 1984) 。无性系变异就象新颖的突变处理方式一样令人感兴趣，它具有产生重要遗传变异的潜在能力。但无性系变异与其它的突变方式一样具有某种有害性，特别是在一个克隆中可能同时发生大量无法控制的突变事件。在典型栽培品种的快速、大量离体繁殖期间发生的无性系变异也显然是有害的。了解和防止这类现象就需要可以适当地测定遗传变异量的方法。由于无性变异在某种程度上涉及到大范围的DNA重排、转座子移动或影响超变区的事件，因此采用适当的探针和RFLP 技术已成为分析这些变化的有效方法。

10.8.7 成本效益分析

分子标记广泛应用的最重要限制因素是成本。由于分子标记技术在人类遗传学和生物学研究的其它领域的广泛应用，所涉及的技术过程已易于实现，熟练的实验室技术员已很容易掌握。但是，目前的 RFLP 方法需要利用昂贵的材料（限制性酶和放射性标记物）和

精细的实验操作。过去曾估计过在不同育种应用中 RFLP 分析的成本 (Beckmann和Soller 1983)。从那以后，已有许多技术的发展使成本大大降低了。这些技术包括可以反复使用的杂交膜以及不需要 CsCl梯度离心纯化的从植物中提取 DNA的新技术。生物素 (Leary等1983)或酶标记探针(Renz和Kurz 1984)技术的使用，可避免在实验室中利用放射性材料。

从植物遗传改良中分子标记应用的成本/效益考虑，最大的问题与Jeffreys等(1985)为重复、高变区发展的探针的可用性、种类及独特性有关。考虑到在真核基因组中存在大量的非编码区，且大量 DNA 占据着这些区域，似乎可以期望一个重复、超变区的小“王国”也存在于 DNA“宇宙”内的其它地方，等待着我们去开发和利用。TRS 等新型分子标记特别适合于 RFLP 难以检测的高度重复区，因而可以获得足够的多样性，以为数较少的探针来覆盖整个基因组。

植物育种家已经在提高作物生产力方面取得了巨大的成功。过50去年中，世界粮食产量的增长，50%以上归功于品种的改良。几乎没有人怀凝生物技术对未来的作物改良会产生巨大的影响。生物技术的某些方面（如抗病虫基因工程）将开拓育种家的工作领域，产生一些过去从未有过的性状。一旦分子标记技术与现有的育种程序相结合，育种家就能以前所未有的速度和精度来鉴定、转移、整合基因。由于分子标记技术直接影响育种方法，它将成为第一个参与商业和政府育种计划的生物技术，也可能是第一个影响植物生产的生物技术。现已能将复杂性状或多基因性状分解为几个遗传成分，研究者有可能在特定的遗传背景下确定每个成分的作用和综合效应。通过检测连锁的分子标记，多基因和主基因的遗传转移将变得更加容易。因此，我们有可能育成综合许多性状的作物品种，这在过去是不可能的。分子标记技术也可能加速从野生种中获取重要基因的进程，这些基因能提高产量、抗性或适应逆境的能力。用同套探针构建分子图谱能在亲缘作物种间建立遗传关系。

几个不同植物群体的研究结果充分证实了在鉴别和定位 QTL、研究与杂种优势和其它遗传性状有关的数量变异方面，标记技术的有效性。当 QTL等位基因转移到其它遗传背景之后，在不同环境下的稳定性还有待深入的研究。Tanksley和Hewitt (1988)指出，所鉴别的遗传因子在适当的遗传背景中评价之前，根据标记与数量性状之间的关联建立育种计划可能是危险的。

分子标记，尤其是 RFLP 的发现给植物育种开辟了新的途径。在实际的育种工作中是否应用这些技术依赖于标记与目标基因的连锁程度，节省时间和借助标记选择与直接选择的相对成本。

第 11 章　QTL分离克隆的战略

前文已叙述了应用分子标记定位数量基因的各种方法及其原理。然而分子数量遗传学的目标正如分子遗传学的一般目标一样，是要分离和克隆数量性状的基因即 QTL，进而研究数量性状基因的结构和功能，乃至最终应用基因工程的手段来操纵和改造 QTL。现在看来，虽然这尚是一个遥远的目标，然而随着基因克隆技术的进展和基因组研究的深入，克隆QTL 已经不是一个可望而不可及的目标了。本章将在讨论一般植物基因分离方法的基础上进一步探讨 QTL 分离与克隆的方法。

11.1　植物基因分离的一般方法

所有的基因分离技术都或多或少地利用了有关基因的下述四个特征：基因具有确定的初级结构（序列），在基因组中占据特定的位置，编码具有特定表达模式的 RNA，大多数基因都具有某种功能。尽管一些技术只适用于一种或少数几种植物，但也有一些可用于从任何植物中分离基因。但是，包括几乎全部经济植物在内的显花植物，只进化了大约一亿五千万年。因此，从某一显花植物所分离的基因一般都可通过异源杂交用于分离其它植物的相应基因。而且当这些基因在导入到亲缘关系较远的物种之后，它们的启动子仍能保留正常的表达模式，因此从拟南芥之类的非栽培植物分离的基因可用于操纵栽培植物。有鉴于此，分离植物基因的重要措施在于发展拟南芥、玉米和番茄等模式植物的先进的遗传学方法、饱和的遗传图谱以及大量的突变体。

11.1.1　以基因功能为基础的克隆方法

以基因功能为基础、应用最为广泛的分离基因的方法包括蛋白质的纯化和突变表型的互补。然而对于许多基因来说，根据其相应的蛋白质性质来克隆基因的主要限制在于，作为基因产物的蛋白质是未知的，难以纯化到足够数量以用于氨基酸的序列分析或相应抗体的制备。

但是，许多植物基因可根据其在细菌和酵母中对突变的遗传缺陷互补的能力来分离。在最近的一个例子中，酵母的八个营养突变体在功能上均能被拟南芥的 cDNA 文库中的克隆互补。又如，一种经遗传工程改造的酵母菌株，可在表型上识别蔗糖载体的活性，现已用它来分离菠菜的编码蔗糖载体的 cDNA 。通过异源表达对几种内膜蛋白质的克隆进一步证实了：植物和酵母利用相似的结构域将蛋白质导向各种细胞器。因此，微生物的互补有希望成为分离编码植物和酵母共有功能的许多基因的有效工具。

尽管有可能通过培养来转化和产生大量的植物细胞，但迄今未能通过对植物突变的互补来实现基因的分离。这一方法的可行性已通过从转化植株中拯救导入的抗菌素抗性标记而得到证实，但由于在细胞培养中表达的突变相对较少，因而这一方法的应用可能会受到

限制。

随着功能尚未确定的已克隆基因数量的增加，有可能通过创造可表达已克隆基因的反义RNA的转基因植株来阐明基因的功能。所谓反义RNA技术就是将某个目标基因的cDNA反向置于表达的启动子之后，导入到植物基因组之中，通过反义RNA的表达抑制或降低该目标基因在植株中的自然表达。从这个意义上看，反义技术的利用等价于创造特定的突变。表达反义结构的的转基因植株已用于鉴别与番茄果实成熟有关的两个基因（Gray 等1992）。

11.1.2 以mRNA或蛋白质表达水平为基础的克隆方法

几种分离植物基因的技术都利用了许多基因具有特征性的表达模式这一优点。许多具有潜在利用价值的植物基因，如涉及有价值的次生代谢物合成的基因，都是在专化组织中表达的。富集这类基因的一种通用方法就是进行差异性筛选。即：采用特定的环境条件（如干旱）之类的指标将不同的植株分开，然后从其组织中或在其发育的不同时期制备mRNA；再用每一mRNA样本产生的标记的cDNA与适当生物的DNA文库序列进行探针杂交，凡能被某种mRNA样本的cDNA以较高程度标记的克隆就可能含有在两种样本间差异性表达的基因。这一方法的主要限制在于它只能用于鉴别那些表达程度相对较高的基因。利用PCR扩增稀有cDNA的新的“差异性显示”技术，可以鉴别低丰度的差异性表达基因（图11.1）。利用一套结构式为5′-$(dT)_{11}dN_2$-3′（dT代表胸腺嘧啶核苷酸，dN代表四种脱氧核苷酸中的任一种）的12种可能寡核苷酸及随机排列的10寡核苷酸作为PCR引物，从第一条cDNA链可扩增出一组cDNA片段。用DNA测序胶比较不同来源（包括生长在不同环境下的不同组织、植株等）mRNA的反应产物，可揭示出不同组织类型之间基因表达的差异。然后利用同样的引物来扩增从胶上所切下的带，以备后续克隆或用作探针来筛选cDNA文库或基因组DNA文库。

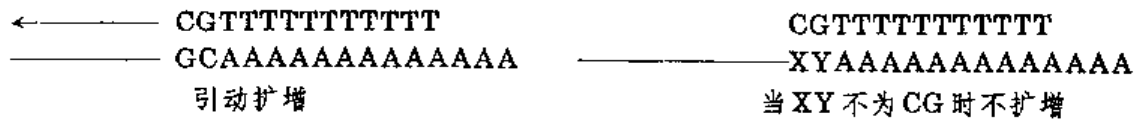
通过上述方法鉴别特定基因的一个问题在于许多基因具有类似的表达模式。因此，为了将这种寻找工作集中于单个基因，需要一些额外的指标，如构建可表达反义结构的转基因植株等。

11.1.3 以DNA插入为基础的克隆方法

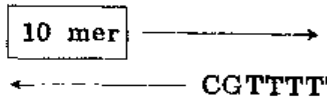
某些种类的突变可较大地改变染色体的结构，已有几种方法以此为基础来分离有关的基因。最富有成效的方法是利用转座子或T－DNA（来自Ti质粒的被转移DNA）示踪将外源DNA插入或靠近目标基因，从而产生可以“标记”该基因的突变。然后利用示踪DNA作探针与被示踪品系的基因组文库杂交，很容易将包围分子示踪物的染色体区域分离出来。转座子示踪主要限于二倍体自花授粉物种如金鱼草和玉米，它们含有特征明确的内源转座因子，并被证明是非常有用的基因分离方法。用这种方法所分离的基因包括与玉米种子发育有关的 *Viviparious－1* 基因和编码转录调节子的 *Opaque－2* 以及与金鱼草花发育有关的几个基因。为了将转座子示踪应用于其它物种，通过转化已将玉米Ac因子的不同衍生物导入到几种植物中，并已发现了转座现象（Grevelding等1992）。这些系统的发展将显著拓展这一方法的应用范围。

1. 从A和B中分离多聚A mRNA，将这两个样本平行用于以下反应

2. 利用序列为T(11)XY可能的12个寡核苷酸之一进行反转录，本例中寡核苷酸为T(11)GC

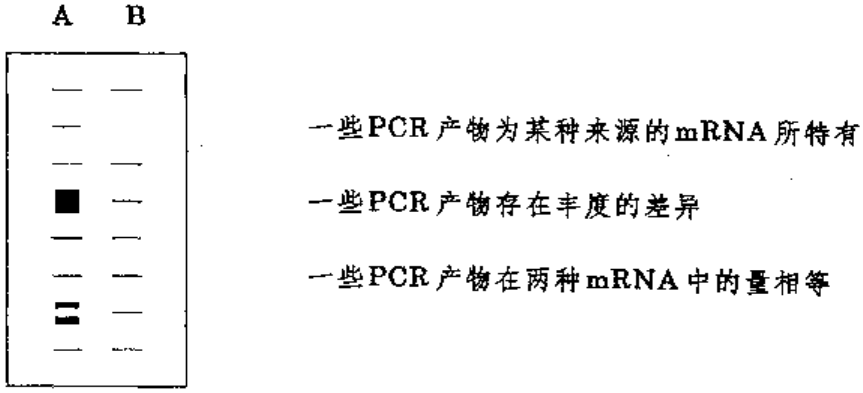


3. 在标记的 PCR 反应中，采用随机10 核苷酸引物和用于反转录的寡核苷酸扩增反转录的 DNA



只有一小部分DNA分子被随机10 核苷酸分析所扩增

4. 利用高分辨率的聚丙烯酰胺凝胶电泳分离 PCR 产物



5. 为了克隆差异性表达的基因，切下谱带，利用PCR反应和在第3步用过的寡核苷酸重新扩增，将扩增产物与载体连接，并转化细菌

图11.1　利用差异显示技术克隆基因

T-DNA 示踪的依据是在由致瘤农杆菌介导的转化产生的转基因植株中，常有一个或多个拷贝的 T-DNA 随机地嵌入基因组中。由于在大多数植物转化期间所采用的组织培养程序本身还会导致突变，因此这一方法对拟南芥是非常实用的，并已发展了大规模转化完整植株的方法 。采用这种转化程序，经 T-DNA 插入可示踪35～40%的突变。由于拟南芥相对来说缺少非编码 DNA，这意味着较高比例（约19%）的转化体具有可见的表型，因此拟南芥对于 T-DNA 示踪是特别有用的。现已产生了10000多个独立转化的拟南芥品系，其中大约一半来自于美国俄亥俄州立大学的拟南芥资源中心。这一方法的主要限制在于：为了生产和繁殖大量完整的转化植株需要付出极大的努力。如果在可供利用的资源中不存在特定的突变，则必须采用其它方法来克隆目标基因。

基因示踪方法的一个重要发展就是利用含有报告基因或强的转录增强子的T－DNA来转化大量的原生质体或培养细胞(Hayashi等1992)。利用报告基因的转录活性可检测到报告基因的结构在启动子下游的异位插入。这一捕获增强子方法的效率取决于诱导培养细胞中

基因表达所需的特定条件的选择。类似地，将强的转录增强子插入到其它处于静止态的基因 (Quiescent gene) 的附近可以诱导基因的非正常表达，这一方法成功的关键在于根据基因活性设计选择的程序。在证实这一方法的潜在可能性的实验中，利用含有 CaMV 35S 增强子的 Ti质粒转化烟草的数百万原生质体，然后在不加植物生长素的条件下对其在培养基上生长能力进行选择 (Hayashi等1992)，成活的组织块显示出过去尚未鉴别、可能涉及生长素合成的基因由增强子介导的表达。

11.1.4 减法杂交 (subtractive hybridization)

减法克隆与基因示踪在概念上是互补的。这一技术的关键步骤是将含有缺失突变的品系的总基因组DNA 与来自野生型植株的高度片段化 DNA进行过量杂交，凡不与突变体DNA杂交的野生型 DNA片段可能就是对应的缺失区域，通过分离拟南芥与赤霉素生物合成有关的基因(*gal*)已经证实了这一方法的可行性(Sun等1992)。这一技术还可以在 cDNA水平上用于分离在某种突变的植株中不表达的基因，或用于分离在自然条件下差异性表达的基因。这种方法的一个成功的实例就是利用减法杂交来分离冰草的盐协迫诱导基因 (Gulick 和 Dvorak 1990)。

与大多数其它方法相比，减法克隆最适用于基因组较小的植物。该技术的关键在于发展可产生较小缺失突变群体的可靠方法；此外，一旦分离到突变，就需要进行深入的遗传分析以确定突变是由于缺失还是由于某种其它的畸变造成的。

上述这些基因克隆方法的应用大大地拓展了用于植物基因工程的基因源，使得植物基因工程突破了仅仅依赖于微生物和病毒来源的基因的局限性。基因工程已为植物育种家展示了改良作物性状的巨大潜力，可望改良的性状包括作物对病虫害和环境逆境的抗性和耐性、产量、品质及植株形态结构等。此外，鉴于生产有用产物的许多野生植物不宜大规模人工繁殖，试图将其有关合成能力的基因导入到已适合于现代农业的作物中已成为植物基因工程的新目标。即使对于复杂的生理性状，只要已鉴别和分离出有关的基因，也可以采用这种方式进行改良。比如将磷酸甘油酰基转移酶基因从耐寒物种转移到敏感型物种之后，由于导入酶的不同底物专一性改变了膜脂组成，从而提高了转基因植株在低温下的生活力 (Murata等1992)。

许多潜在而有益的改良都涉及到改变生物合成途径中有关基因的表达，从而控制理想的和非理想的化合物的数量。例如，在甘蓝型油菜 (*Brassica napus*) 中利用反义构型的硬脂酰ACP 脱氢酶基因的表达可以导致种子油分中硬脂酸含量的显著增加 (Knutzon等1992)。在另一些情形下，通过导入编码不同底物专一性或变构特性的酶的基因可能获得有用的改良。例如加州月桂树中的脂酰硫酯酶基因，具有极高的中链脂肪酸专一性，可用于提高拟南芥中链脂肪酸的含量。类似地，将催化淀粉生物合成物中第一个关键反应的 ADP－葡萄糖焦磷酸化酶的反馈不敏感基因导入到马铃薯中显著地增加了马铃薯块茎中累积的淀粉数量 (Stark等1992)。此外，通过改变有关的转录激活子的表达，可能提高植物中有用代谢产物、蛋白质或贮藏产物的水平。最近已证实了这一途径的潜在可能性。编码转录因子的玉米基因已用于改变拟南芥和烟草中的花色素苷生物合成途径表达的数量及组织专一性 (Lloyd等1992)。上述有关的进展以及在克隆重要性状的主基因方面的成

功为数量性状基因 (QTL)的克隆奠定了基础。

11.2 以图谱为基础的基因克隆

如上所述，植物基因工程已取得了长足的进步，然而植物育种的许多目标基因还不是分子生物学的研究对象，问题在于现成的各种分子克隆的方法依然有较大的局限性。前节介绍的克隆基因的部分方法有赖于对其基因产物的了解，而迄今为止如抗病、抗逆及各种与产量或品质有关的基因的作用机制和产物都是未知的。例如从植物基因工程的角度看，抗病基因是最重要的目标基因。尽管经典的遗传分析已判明许多抗病性状都是由单基因决定的显性性状，符合基因工程操作的条件，然而无论是病理学还是生物化学的研究都未能指示出任何一种抗病基因的产物，因此也不可能循 mRNA 和 cDNA 的途径克隆抗病基因，当然真正的植物抗病基因工程也就无从谈起了。虽然转座子示踪法不依赖于对基因产物的了解，且曾被广泛用于微生物基因的克隆，但在植物上的应用受几个因素的限制。首先，该法只适用于转座子可行使转座作用的植物；其次，为了发现插入突变必须筛选多达10万个以上的个体或组织细胞；第三，许多基因决定的性状难以在个体表型或组织水平进行选择，如数量性状的基因。而减法杂交法只适用于由缺失造成的突变基因的克隆。

定位克隆技术 (Positoinal cloning)或以图谱为基础的基因克隆 (Map-based gene cloning) 在分离未知产物的基因方面有更为广阔的应用前景。有时这类有关的方法学也被称之为反向遗传学(Reverse genetics)。该法的基本前提是基因定位，然后以紧密连锁的分子标记为起点，通过染色体步行，逐步向目标基因靠近，最终克隆基因。应用这种方法有可能克隆植物的任何基因，只要该基因的突变型（等位基因）可以在遗传上鉴别。目前在人类中应用定位克隆法已成功地克隆了几个控制遗传病的基因。

定位克隆法包括几个重要的步骤：(1) 将目标基因定位在高密度的分子标记连锁群上；(2) 利用 PFGE 将连锁标记的遗传图距转换成物理距离；(3) 构建 YAC 文库，找到含连锁标记的YAC 克隆，并通过克隆的排序获得含目标基因的 DNA 大片段；(4) 通过转化和功能互补试验证实基因所在的DNA 片段(图11.2)。随着高密度分子遗传图谱在一些植物中的陆续建立，以图谱为基础克隆目标基因已成为现实。

11.2.1 利用紧密连锁的分子标记跟踪基因

尽管现在已将大量的目的基因定位到了分子标记图谱上，但它们与标记之间的连锁并不是非常紧密，一般大于2.5cM。这样的距离应用于经典的植物育种已经足够了，但还不足以通过遗传连锁和染色体步行来克隆基因。即使是象拟南芥这样的模式生物，平均而言1cM 相当于150kb。在一些重要的栽培植物中，这种情形就更糟，它们的基因组要大得多，即使如番茄和马铃薯等基因组相对较小的物种，1cM至少相当于500～1000kb。而在玉米、小麦和大麦中，1cM则相当于2000～3500kb甚至更高。

通过遗传连锁进行基因克隆有赖于将目标基因定位到标记间紧密连锁的饱和分子图谱上。例如，对于1400cM 的番茄基因组来说，为了有90% 的把握每隔2cM 有一个分子标记，就必须在图谱上定位大约1750个标记。为此需要花费极大的努力，也许需要几年的时间才

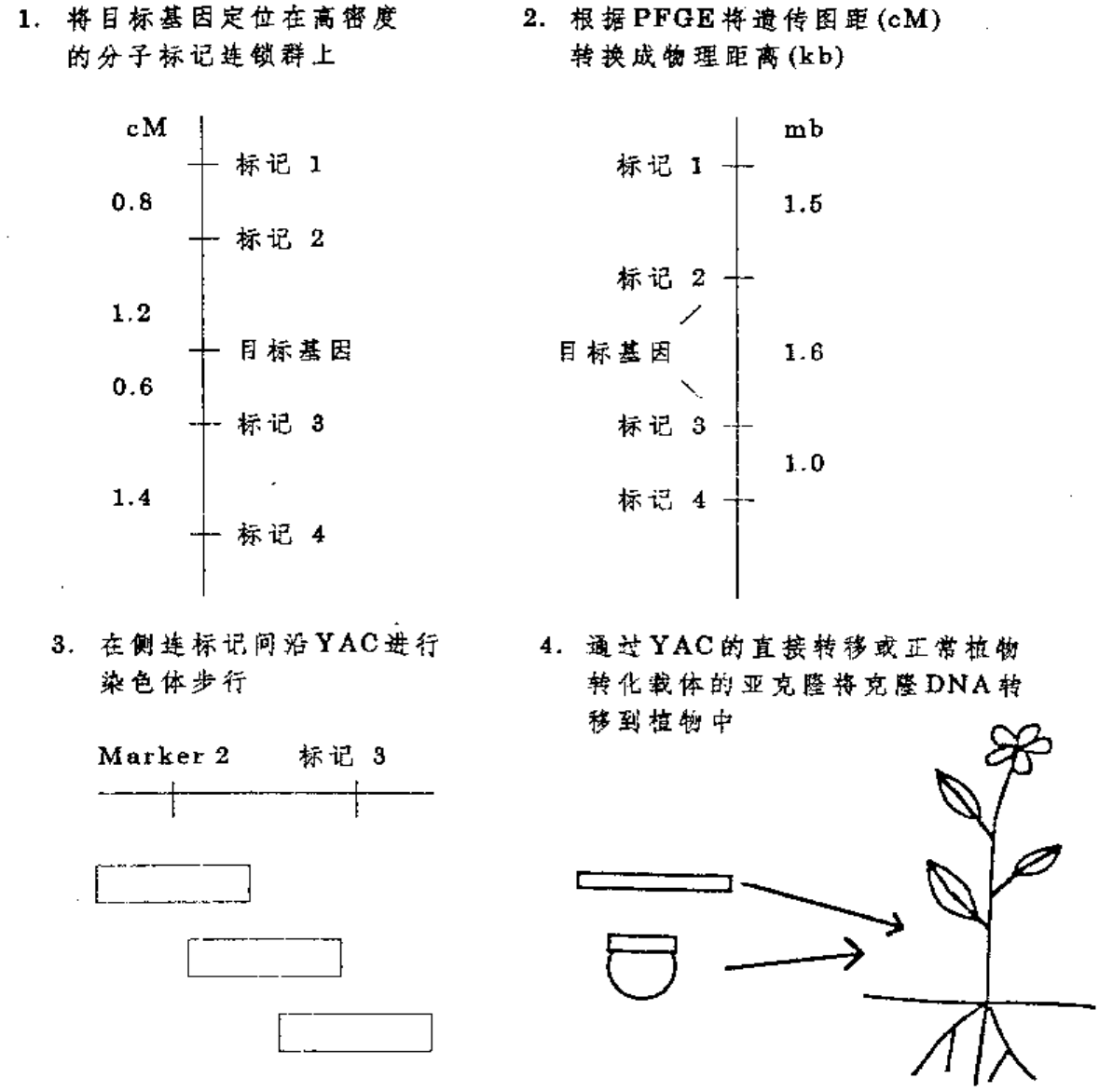


图11.2　以图谱为基础克隆基因的步骤

能实现。近几年为加快这一进程，已发展了几种方法来快速定位基因。首先，可以利用三体、单体、代换或附加系将单拷贝的 RFLP 标记定位到染色体上(Helentjaris 等 1986a; Young等1987)。其次，近等基因系可用于寻找与导入基因紧密连锁的标记(Young等1988)。这两种方法的优点在于一次杂交可以分析一个以上的标记。此外，其它技术诸如个别染色体的分离和随后的特定染色体文库的构建、利用切割后的染色体进行微克隆、以及由原生质体融合获得不对称杂种都成功地用于人类基因组的研究，也同样可能为复杂的植物基因组某特定区域的饱和提供新的方法。已定位到特定区域的标记还需要在包括几百个个体的大群体中对目标基因进行经典的分子标记分析，并构建高分辨率的遗传图谱。

即使构建了有大量分子标记聚集在目标基因附近的高分辨率的遗传图谱，我们仍然不清楚这些标记之间相互间隔多少物理距离(碱基对)。 这是因为遗传图谱和物理图谱只是

在标记排列的线性顺序上是相同的。尽管从基因组的大小和遗传图谱的大小可以推测 1cM 的图距相当于多少物理距离，但这决不是意味着在染色体的某特定区域该值是固定不变的。通常的情形是，染色体着丝粒附近的区域重组受到抑制，而其它区域则具有较高的重组率。这种差异很容易使图距偏离其平均数一个数量级(Ganal等1989)。而且，遗传图谱的分辨率还受到可用于分析的个体的限制。通常要研究几百个以上的个体(包括子代测验)不是一件容易的事，因此遗传分辨率很少低于1cM。而另一方面，这1cM 的图距对于传统的分子生物学技术已显太大。但是，最近大片段DNA 技术的发展已缩小了遗传图谱和物理图谱之间在分辨率上存在的缺口。各种技术的分辨范围如图11.3所示。

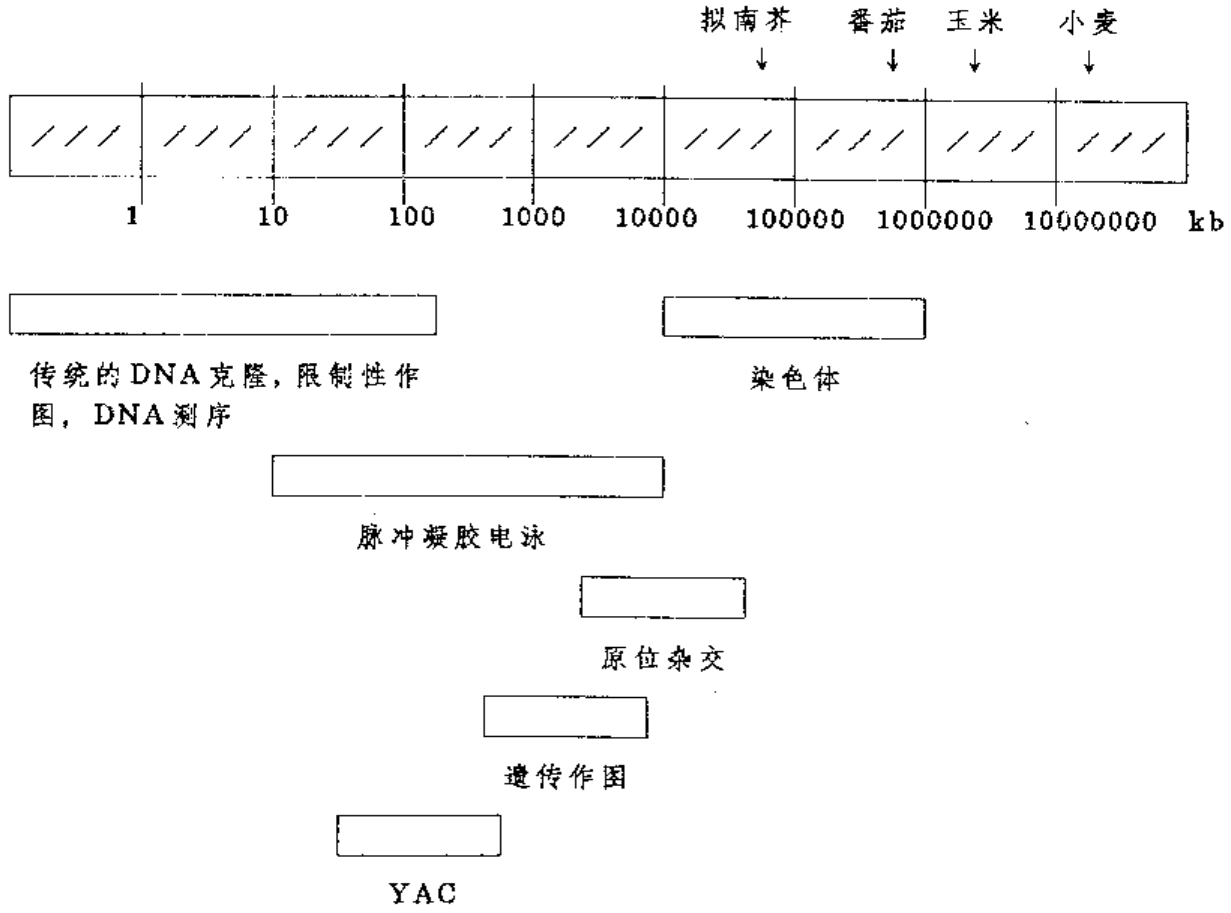


图11.3　分子生物学、遗传作图和细胞遗传学技术的局限性的比较

11.2.2　利用脉冲凝胶电泳进行大尺度物理作图

在脉冲凝胶电泳 (PFGE) 发展之前，DNA 分子的最大分辨率未超过100kb。PFGE技术可以分辨约5～10Mb (百万碱基对)的 DNA分子。这种分辨率的提高基本上是通过周期性地改变电泳时电场的方向来实现的。利用这一特征已发展了不同的工作系统 (Schwartz和Cantor 1984; Carle等1986; Chu等1986)。所有这些不同的系统都将 DNA 分子的分辨率达到了百万碱基对的水平，使之完全适合于遗传图谱的工作范围。

由于通常的DNA 分析技术不能产生 PFGE 分析所需的大 DNA分子，因此另一个技术进步的重要步骤是分离这种大分子的 DNA。 Jackson和Cook(1985) 成功地发展了用固相支持物(凝胶)分离 DNA的方法。将细胞包埋在凝胶中能有效地防止 DNA分子的剪切，使得从许多生物中分离染色体大小的 DNA分子成为可能。由于植物细胞被细胞壁所包围，由此造成的问题直到最近才得以解决。 因此现在植物的 DNA分离技术在凝胶包埋之前包括了原生质体制备的步骤。 为了防止由于内源核酸酶对 DNA 的降解，原生质体在凝胶中的解体是在高浓度的EDTA 中附加去污剂(如 SDS 等) 的条件下进行的，并且细胞内的蛋白质是用蛋白酶降解的。

但是，植物染色体的平均大小超过了 PFGE 所能分离的范围的上限。例如，即使是染色体相对较小的番茄，其大小也在33～64Mb。因此这类 DNA 必须用限制性酶切割以产生适合 PFGE 分离的片段。采用识别6碱基的大多数限制性酶切割所产生的 DNA 片段太小(平均5～50kb)，不适合于用 PFGE 分离，因此可以用于 PFGE 切割的限制性酶的数量是极为有限的，它们通常具有下列两个特征之一：

(1) 识别位点具有6个以上碱基对，几乎完全由 G或C 组成(植物基因组的GC 平均含量只有40%)。这类酶在基因组中识别位点较少，因而能产生更大的 DNA 片段，如 *Not* I (GCGGCCGC)和 *Sfi* I(GGCCNNNNGGCC)。

(2) 识别位点含有在大多数真核生物基因组中非常稀有的碱基序列。这些酶的识别序列通常具有几个 CG序列的碱基对，这类位点在复杂基因组， 包括大多数植物中是普遍缺乏的。此外，这类CG 序列是植物DNA 甲基化的目标，而甲基化则能防止大多数限制性酶切割这样的序列。 这类限制性酶 如 *Bss* HII (GCGCGC)、*Mlu* I (ACGCGT)、*Nru* I (TCGCGA)、 *Sac* II(CCGCGG) 和 *Sma* I(CCCGGG)。

尽管考虑到上述这些特征，在许多植物中仍然找不到可以产生几百万碱基大小片段的 DNA 限制性酶，因此需要在 PFGE 凝胶上获得大于 2.5Mb的片段是很难的 (Ganal 和 Tanksley 1989)。还有一些植物的基因组具有相对较高的GC 含量和相对较低的甲基化，因此真正的问题在于如何找到切割位点稀少的限制性酶。

一旦 DNA被含稀有切割位点的限制性酶切割，并经 PFGE 分离，就可以转移到滤膜上，然后象从标准凝胶上转过来一样，用于分子杂交。为建立大尺度物理图谱，这类转移的印迹应与紧密连锁的分子标记杂交。凡是在至少两种酶酶切均能与同一片段杂交的那些标记被认为在物理上是相互连在一起的。利用双酶切可以相当精确地确定其距离。这种图谱的构建与一般的限制性图谱一样。

PFGE 技术已被成功地用于分析番茄抗烟草花叶病毒的基因 *Tm-2a* 所在的区域 (Ganal 等1989) 和编码不同植物核糖体 RNA 的基因 (Ellis等1988) 等。 PFGE 的另一个令人感兴趣的应用是从甜菜分离抗线虫的染色体片段 (Jung等1990)。

11.2.3 酵母人工染色体大DNA片段的克隆

一旦在目标基因两侧发现与之紧密连锁的分子标记，并且为确定目标区域的边界已利用 PFGE 技术构建了物理图谱，那么就必须克隆这些标记之间的整个区域，以鉴别和分离目标基因。如上所述，两个侧连标记通常相隔数百kb的距离。利用 *E. coli* 所进行的传统

克隆只能分离小于50kb的DNA 片段。 因此利用这类标准载体通过重叠克隆(染色体步行)克隆这样大的区域是不可行的。 现已发展了可以满足这样大容量需要的克隆系统。 与许多其它的克隆系统相反，该克隆系统是以真核的宿主酵母为基础的 (Burke等1987)，因为在酵母染色体上所有必须的元件都是已知的并已分离出来。酵母人工染色体载体包括着丝粒、两个端粒、自主复制序列及选择性标记，这些成分已足以维持在酵母中的稳定复制，它们已被组装到约10kb的载体上。另一个优点是，由于利用真核宿主，该系统较少出现克隆的选择偏差。

通常采用在 PFGE 分析中分离高分子量 DNA 的方法来分离用于YAC 克隆的DNA，并用 *EcoR* I 部分酶解或在连接之前进行物理切割，然后将连接产物转化到酵母原生质球中，并按照每一载体臂上的标记选择转化体，最后将转化体挑选出来进行保存。对这种 YAC库的筛选可通过菌落杂交或利用 PCR 技术进行。YAC 似乎具有相当大的插入容量，可大于1000kb。但平均大小在100～300kb。

过去所构建的大多数YAC 库都是动物基因组的。但在一些植物中如拟南芥等现已构建了完整的YAC库。近年来发现在已构建的YAC库中，嵌合YAC 的比例相当高，有的高达40～60%。所谓嵌合YAC 是指 YAC插入片段中包含了两个或更多的在染色体上不连续的 DNA片段，已证明它主要是由于不同 YAC在同一酵母细胞中发生同源重组的结果。嵌合 YAC的存在对YAC 重叠群构建和大尺度物理图谱分析都是不利的，有时结果不甚明确。因此，在努力改进YAC 系统的同时，又发展了人工细菌染色体(Bacterial Artificial Chromosome, BAC)，以期构建大基因组的 BAC 库。BAC 库的插入 DNA的容量较小，一般为100kb左右。此外，也有实验室采用 P1 克隆系统 (PACs) 或非洲粟酒酵母 (*Schizosaccharomyces* Pombe) 克隆系统——非洲粟酒人工染色体(Pombe Artificial Chromosome)，也名之为 PACs。

11.2.4 大片段DNA分子的转移

一旦通过染色体步行从特定植物克隆得到含有目标基因的区域，就必须进一步确定目标基因在该区域内所在的具体位置。对于大多数植物来说，将DNA 转化到突变体中或在抗病情况下转化到感病植株中就可以做到。迄今还没有将整个 YAC 从酵母转移到植物中的报道，尽管在动物中已有几个报道 (Pachnis等1990)。但是， 另一项可行的技术是将 YAC亚克隆到 cos质粒转化载体中。而在尚无有效转化系统的植物中，这要复杂得多。对这一问题的可能解决办法，读者可以参考人类囊状纤维变性基因的克隆 (Rommens等1989)。

11.2.5 实例

美国康乃尔大学 Tanksley实验室1993 年底在美国《科学》杂志上报道了第一例以图谱为基础的基因克隆在植物上的应用 (Martin等1993)。我们以此为例加以介绍。

已知番茄对细菌性茎腐病 (*Pseudomonas syringae* pv *tomato*) 的抗性是由第5染色体上的 *Pto* 座位控制的。 通过抗病纯合体 *Pto*/*Pto* 与其感病的近等基因系 *pto*/*pto* 杂交获得 F_2群体， 对其中251个 F_2个体进行 RFLP 分析和抗性鉴定，发现了与抗性基因

Pto 共分离的 RFLP 标记 TG538。然后用此标记去筛选番茄的YAC 文库，鉴别出与该标记杂交、长 400kb的克隆PTY538-1。通过反向PCR 分离出与 PTY538-1 左臂和右臂相对应的两个末端特异性探针 PTY538-1R 和 PTY538-1L，并将其定位于高分辨率的连锁图谱上(图11.4A)。其中 PTY538-1L距离 *Pto* 1.8cM，而 PTY538-1R 与 *Pto* 共分离。为了证实 PTY538-1包含有 *Pto*，就必须鉴别出在 PTY538-1R 和 *Pto* 之间发生重组的植株。为此，利用标记 TG538和 PTY538-1R 分析了来自不同 F_2群体、F_3家系和 50 多份品种的总共 1300个植株。从中鉴别出一株既是与 *Pto* 关联的 TG538等位位点的纯合体又是与 *pto* 关联的 PTY538-1R 等位位点的纯合体。来自这一植株的所有后代均抗茎腐病，表明该植株为 *Pto*/*Pto* 纯合体。这一结果说明 PTY538-1 包含了 *Pto* 座位。

用脉冲凝胶电泳分离出来自PTY538-1的DNA，并用于与叶片 cDNA文库的大约920,000个菌落进行探针杂交。用PCR 扩增克隆中的cDNA 插入序列，并用于与在 *Pto* 区域发生重组事件的85个植株进行探针杂交。其中两个克隆 CD127和 CD146(均长1.2kb)含有相互交叉杂交的序列，且 CD127与 *Pto* 共分离。CD127代表了一个与 *Pto* 相关的基因家族，该基因家族中至少 6个成员在遗传上与 *Pto* 共分离。为利用遗传互补测验确定 CD127基因家族是否抗茎腐病，选择代表两类大小不同转录产物的 cDNA 克隆 CD127 (1.2kb)和 CD186(2.4kb)用于分析。转化试验表明对番茄茎腐病的抗性是由于导入 CD186 插入序列的结果。在以转基因植株作亲本所产生的回交群体中 CD186与抗性基因的共分离得到了证实。最后确定了总长2.4kb的CD186插入序列的 DNA 顺序，并发现在靠近35S CaMV启

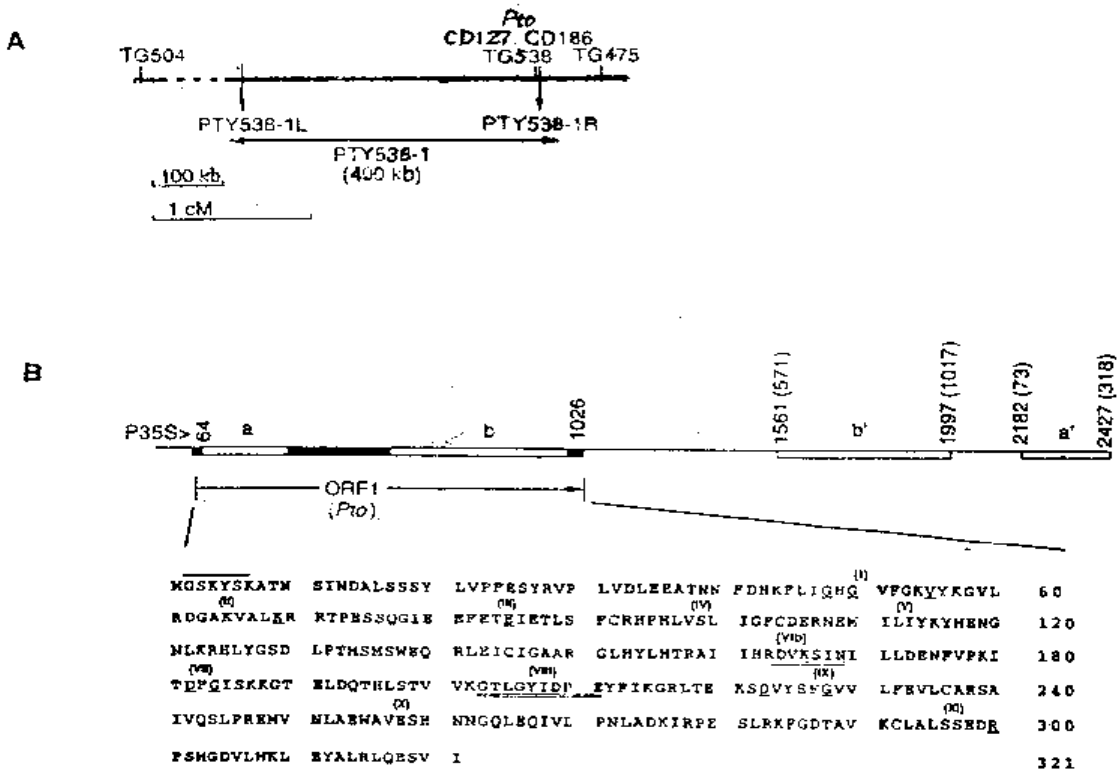


图 11.4 番茄第5染色体 *Pto* 附近的遗传图谱(A)和CD186 cDNA插入序列的物理图谱(B)

动子的区域有一个963bp的开放阅读框(ORF1)(见图11.4B)。由它编码的蛋白质氨基酸序列类似于丝氨酸-苏氨酸蛋白质激酶。

最近，美国加州大学的Staskawicz实验室和哈佛医学院的Ausubel实验室同时应用以图谱为基础的克隆方法分别克隆了拟南芥的抗细菌病害的基因 RPS2，发现该基因是富含亮氨酸的跨膜蛋白 (Bent 等1994; Mindrinos等 1994)。高密度的分子作图和 PFGE 及YAC 等大片段DNA 技术的结合已经打开了通向栽培植物基因克隆和分析的大门，使得植物抗病基因的研究登上了分子生物学舞台。不难设想，定位在分子连锁图谱上的 QTL，也能应用类似的方法予以克隆。可以预言首先成功的将是决定数量性状的主效基因或具有较大效应的 QTL，因为这类基因的作用比较容易通过转化的方法进行求证。如何证实微效 QTL 的作用，无疑是对分子遗传学家的一个新的挑战。

11.3 随机cDNA 序列分析的应用

这是一种在基因组中全面分析所有表达基因序列的方法，也是基因组研究的一个新的战略途径。最近自动化测序和机器人技术的改进使得大规模的 DNA 测序变得现实可行。装备2台自动 DNA测序仪的一家实验室每天可从70个 cDNA 克隆中的每一个获得约500个碱基的序列。简单地挑选随机cDNA 克隆，获得其5′末端的部分序列，并将该部分 cDNA 序列的6种可能的翻译产物与不同数据库中已知蛋白质的序列进行比较，已给5000 个人类克隆中的约18% 确定了可能的功能 (Adams等1991) 。类似地，对首批1000 个水稻的部分序列和拟南芥的数千个序列进行分析，已给其中8% 和20% 的序列确定了可能的功能，其分析大多是根据其与非植物基因的同源性进行的。鉴于这种方法可用于将植物生物学与其它生物的大量序列信息联系起来，因此，人们期望能获得大量的部分序列。以美国的基因研究所为代表的几个研究所每天已可分析1000个以上的 cDNA。因此，在九十年代结束的时候，就可能完成拟南芥和水稻的2万到4万个不同的cDNA 序列的部分序列分析工作。

目前对大多数目的基因都无法仅仅根据其核苷酸序列来鉴别。可鉴别其功能的一种方法就是将其定位在遗传图谱上以便分析其与位于同一位置的突变事件的相关性。在拟南芥中，将克隆与YAC 文库杂交已定位了许多 cDNA。将cDNA 与YAC 杂交，将最终把所有的 YAC 定位到遗传图谱上。根据若干 YAC 克隆与同一 cDNA 的杂交结果就可以鉴别出覆盖同一基因组区域的YAC 克隆之间的跨叠。通过与已定位的 RFLP 标记进行杂交可能将这类彼此邻接的克隆或称为克隆重迭群 (Contig) 定位到遗传图谱上。在理论上，将约2000 个cDNA 克隆与现有的 YAC库杂交就足以产生整个拟南芥基因组的一套排序YAC。不难想象，在大量的随机 cDNA 序列中，不仅包含了许多质量性状的基因，也包含了许多数量性状的基因。问题在于如何判断那些编码序列的功能，这是分子遗传学家面临的又一个挑战。长期以来，遗传学总是从性状的判别入手，来鉴别、分离和克隆基因。 从已知基因序列来分析基因的功能则是现代反向遗传学的主要研究课题。

改进现有的测序技术包括进一步提高测序的效率和可靠性是采用随机 cDNA 测序方法克隆QTL 的先决条件。目前的测序技术均以化学法或酶法为原理。现有的技术能力和测序的工作量之间的尖锐矛盾表明，应在原理上探讨全新的测序技术。科学家们正试图彻底

抛开 Sanger原理，应用或结合物理学的方法，进行 DNA 测序的第二次革命。最为可能的下一代测序技术有质谱术、荧光分光光度术、核磁共振术和电子显微术等，它们都是灵敏的物理方法，这些方法可以彼此或与更为常规的生化分析方法相结合。据悉，美国 Los Alamoss Natioanl Laboratory 和Life Technologies, Inc签定了为期三年的合作研究协议，计划发展一项新的测序技术，实现一次完成50kb DNA 的序列分析，使分析速度达到 1000bp/秒，主要原理是以四种荧光染料标记的四种核苷酸合成DNA单链，借助DNA外切酶的作用进行碱基连续释放，同时进行激光识别和记录。大规模的DNA 测序还处在设备和路线的建立与完善阶段。现在已在人类的 X 和 21 染色体经克隆 YAC- cosmid-plasmid 路线用自动测序仪以随机测序方式进行大规模测序的试验。到目前为止，已测序人类基因组 DNA总长度约为 35,588,275bp (占人类基因组的1.1%)，其中 cDNA 约占 20%。美国 Washington 大学的Wilson等(1994)最近完成了迄今最大的DNA 片段，总长2.1 Mb (2100kb)的秀丽隐杆线虫第3染色体DNA全序列测定。在新的五年计划中，要求再完成几个Mb 长度的人基因组DNA测序。美国TIGR (The Institute of Genomic Research) 估计人类基因的总数约 65,000个。到目前为止，已在基因组中定位的基因座位已有31,048个，已定位的基因有3,216个，已克隆基因 3,103个(约3%)，已测序(包括部分测序)基因1,495个 (约1.5%)。由于 cDNA代表着基因的编码序列，所以目前已有实验室开始大规模 cDNA 测序，到1994年7月底，cDNA已测序总数为 19,834个 (EST)，以平均长度为300bp计算，总长度达5,950,200。预计在近两三年内将完成全部人类 cDNA的测序。

尽管许多植物的基因组研究已取得了重大进展，但目前有关数量性状的遗传研究还停留在根据连锁遗传的资料进行QTL作图的水平上，只能大体上了解QTL在染色体上的位置(cM)，尚未深入到QTL的克隆与序列分析。分辨率最高的物理图谱——DNA序列的测定，将会导致数量性状遗传改良的深刻变革。作者认为可能的变革包括以下三个方面：

1. 从全基因组出发，了解同一性状的多个QTL 之间的关系。即 QTL 在基因组上的分布、精确位置(bp)及其相互关系；不同基因座位间QTL的调控和表达；影响同一性状的主效和微效基因及其相互作用。

2. 通过 DNA 序列分析，了解单个QTL的遗传机制。即QTL的基因结构(内含子、外显子、调节区等)；同一性状多个 QTL 之间DNA 序列、结构和功能的比较；单个QTL 内部的调控。

3. 对植物数量性状实施真正的遗传操纵。即性状表达的时间和空间；通过遗传工程的方法实现多个QTL 效应的累加和重组以获得前所未有的可遗传变异；重新设计和改变 DNA 的序列实现QTL的人工改造。

11.4 QTL 效应分析与歧化选择

利用以图谱为基础的克隆方法或用随机 cDNA 顺序分析都可以最终克隆数量基因，但正如在前两节所讨论过的那样，克隆数量基因的关键在于如何有效地判断其效应。在前一种克隆方法中，对基因效应的分析，既涉及对 QTL 的分解（即将控制同一数量性状的多个 QTL 分解开来）和定位，又涉及对转化的克隆基因效应的证明。而在后一种克隆方法

中，也要力求建立 cDNA 序列与各种 QTL 的关系。并且最后求证某个 cDNA 的基因效应。另外从长远的和发展的观点看，如能提高 QTL 效应鉴别分析方法的灵敏度和可靠性，在第11.2节介绍的以 DNA 插入为基础的转座子和 T-DNA 示踪克隆法以及以 DNA 缺失为基础的减法杂交克隆法，也都可能用于克隆 QTL。对于主效性的数量性状基因和效应较大的 QTL，在理论上可以直接采用已有的基因克隆方法来克隆单个的 QTL。但当同一性状受多个效应较小的 QTL 控制时，QTL 克隆所面临的问题之一是如何将控制同一性状的多个 QTL分解开来。采用本书第6~9章所介绍的 QTL 作图方法就可以在分解 QTL 的同时确定各 QTL效应的大小及其所在的染色体位置。在此我们介绍一种通过选择来鉴别和分解多个 QTL 的方法。

11.4.1 理论

采用数量遗传的研究方法鉴别多基因效应在很大程度上取决于供试亲本间表型均值的差异，反过来这种差异又在很大程度上取决于亲本间具有相同效应的等位基因(同效基因)的分布 (Jayasekara和Jinks 1976)。如果对于某一性状，两个亲本间存在 k 对基因差异，且各对基因的效应相等，则当所有的增效等位基因（即增加表型值的等位基因）集中（联合）分布于一个亲本，而所有的减效等位基因（降低表型值的等位基因）集中分布于另一亲本，例如两对基因，亲本的基因型分别为 *AABB* 和 *aabb* 时，两个亲本之间的表型差异最大。但是，当一部分增效等位基因和一部分减效等位基因分散分布于两个亲本之间时，它们之间的表型差异将减小，且当增效等位基因和减效等位基因完全分散分布于两个亲本之间，例如当两个亲本的基因型分别为 *AAbb* 和 *aaBB* 时，它们之间就可能不存在可以观察到的表型差异。而事实上在两个亲本间是存在真实遗传差异的。且不难看出，当不同对基因间不存在基因连锁时，基因联合态组合 *AABB*/*aabb* 和基因分散态组合 *AAbb*/*aaBB* 具有遗传组成完全相同的 F_1和 F_2。然而，在育种实践中，育种家通常选择性状差异较大的两个遗传材料作为杂交亲本，并力求双亲表型性状的互补。显然这种传统的亲本选择方法将把那些相互之间存在真实遗传差异而表型近似的遗传材料排除在候选亲本之外。

如果数量性状由多个 QTL所控制，这时具有相同效应的QTL等位基因通常分散分布于多个不同的遗传材料中。Xu(1989)、Xu 和Shen (1992a)在加性-显性模型的基础上，提出了多个QTL同效等位基因在遗传材料间分散分布的三种检测方法，可揭示出表型相似的遗传材料间可能存在的真实遗传差异（基因效应）。这三种方法是：(1)检验遗传材料间F_2的表型方差与环境方差的同质性以及 P_1、P_2、F_1、F_2、B_1 ($F_1\times P_1$) 和B_2 ($F_1\times P_2$) 六个群体间的方差同质性。其中环境方差由3个非分离群体 P_1、P_2和 F_1的合并方差估计，F_2的表型方差与环境方差的同质性采用F测验进行检验，而6 个群体间的方差同质性采用 χ^2测验进行检验。(2)检验上述6个群体表型均值间的差异显著性。在基因效应相等的假定下，基因完全分散态组合的 B_1和 B_2的表型均值相等，且当只存在加性基因的作用时，6个群体的表型均值均相等，而基因联合态组合则相反。(3)对遗传材料间的 F_2群体进行歧化选择(Disruptive selection, Thoday 1972)，获得具有极端表型的两类品系，然后比较这两类极端品系的杂交组合(基因联合态组合)与原始亲本间组合(基因分散态组合)之间遗传参数估值的差异。当由平均值所估算的基因加性效应之和 [*d*] 及基因显性效

应之和 [b]的估计值在两类组合间存在显著差异(例如在分散态组合中 [d] 的估值近似于0，而在联合态组合中与0存在显著差异)时，即表明原始亲本间存在遗传差异。 由于二级统计量不受基因分散分布的影响，如果原始材料间存在真实遗传差异， 则由两类组合所估算的遗传方差的加性分量 D和显性分量 H 是同质的。对极端品系进行遗传分析，还可以获得经典数量遗传分析的几乎全部结果，包括遗传模型的检验、有效因子数和基因联合-分散度的估计以及重组近交系分离极限的预测等， 从而揭示分散于不同遗传材料中的多个 QTL的遗传信息。因此，基因分散态组合的检测及其遗传分析可以发现隐藏在遗传材料中、过去通常被育种家忽视的QTL 有利等位基因。利用这些分散基因的重新组合，可以获得农艺性状超亲的两类变异体。 在 QTL 克隆中，当增效等位基因为有利等位基因时 ，高表型的极端材料可用作QTL 有利基因的供体，而对应的低表型极端材料就可用于验证转数量基因植株的 QTL 效应。

11.4.2 实例

下面以水稻分蘖角度的研究为例，说明数量基因效应的鉴别和歧化选择的应用。水稻的分蘖角度是指水稻抽穗前主茎与其分蘖之间的夹角，它在决定水稻株型及水稻生长发育方面起着重要作用。过去对水稻分蘖角度的遗传研究多以极端的 “匍匐型” 突变体（茎秆匍匐于地面生长的一种突变体）为材料。1985年， 我们在水稻早熟籼型水稻品种 5002和竹菲10的 F_2群体中发现了分蘖角度的超亲变异。经过重新配制组合证实了这两个亲本间虽然表型基本相近(分蘖角分别为 14.9和15.3°) 但存在真实的遗传差异。 后来又发现了另一对类似的亲本 HA79317-7 和珍龙13(分蘖角分别为17.9和12.9°)。为证实亲本间可能存在的多个 QTL 同效等位基因的分散分布，分别以这两对亲本为材料配制杂交组合，并自交产生 F_2。通过歧化选择，从每一组亲本的 F_2群体选择出两种类型的极端品系，一类品系有较大分蘖角(分蘖角为25.7～28.0°，另一类具有较小的分蘖角(4.7～5.5°)。分别利用原始亲本和极端品系作亲本，培育由6个群体 (P_1、P_2、 F_1、F_2、 B_1、B_2) 组成的两套材料，用于检测不同数量基因的分散分布(图11.5)。 经方差同质性检验和遗传参数的估计，发现两组原始亲本中均存在控制分蘖角度的两对基因的差异，其同效等位基因分散分布于原始亲本而集中(联合)分布于所选的极端品系中。控制分蘖角度的数量基因只受基因加性效应影响，而与显性和上位性效应无关 (Xu 和 Shen 1992b)。

为进一步了解在两组原始亲本(4个品系)间与分蘖角度有关的基因的异同， 我们把从两组亲本的分离群体中所获得的各类稳定遗传品系进行相互杂交。 结果表明， 两组亲本间分蘖角度最小的品系相互杂交，其 F_2群体的分离未显著超过其亲本品系，显示来自不同组合、分蘖角度最小的极端品系具有相同的基因组成，它们已集中了所有的使分蘖角度降低的 QTL减效等位基因。分蘖角度适中的品系相互杂交，其中一部分F_2群体的分离显著超过其亲本品系，显示其中一些品系存在 QTL同效等位基因的分散分布。分蘖角度最大的品系相互杂交，其 F_2群体的分离仍显著超过其亲本品系，显示来自不同组合、分蘖角度最大的极端品系间仍可能存在 QTL同效等位基因的分散分布。为证实这一推论，我们从F_2群体中进行第二轮的歧化选择(图11.5)，获得了稳定遗传、分蘖角度超过上一轮极端品系的大分蘖角材料，其分蘖角为55.9～56.5°。而所获得的分蘖角度较小的极端品系， 其表现型

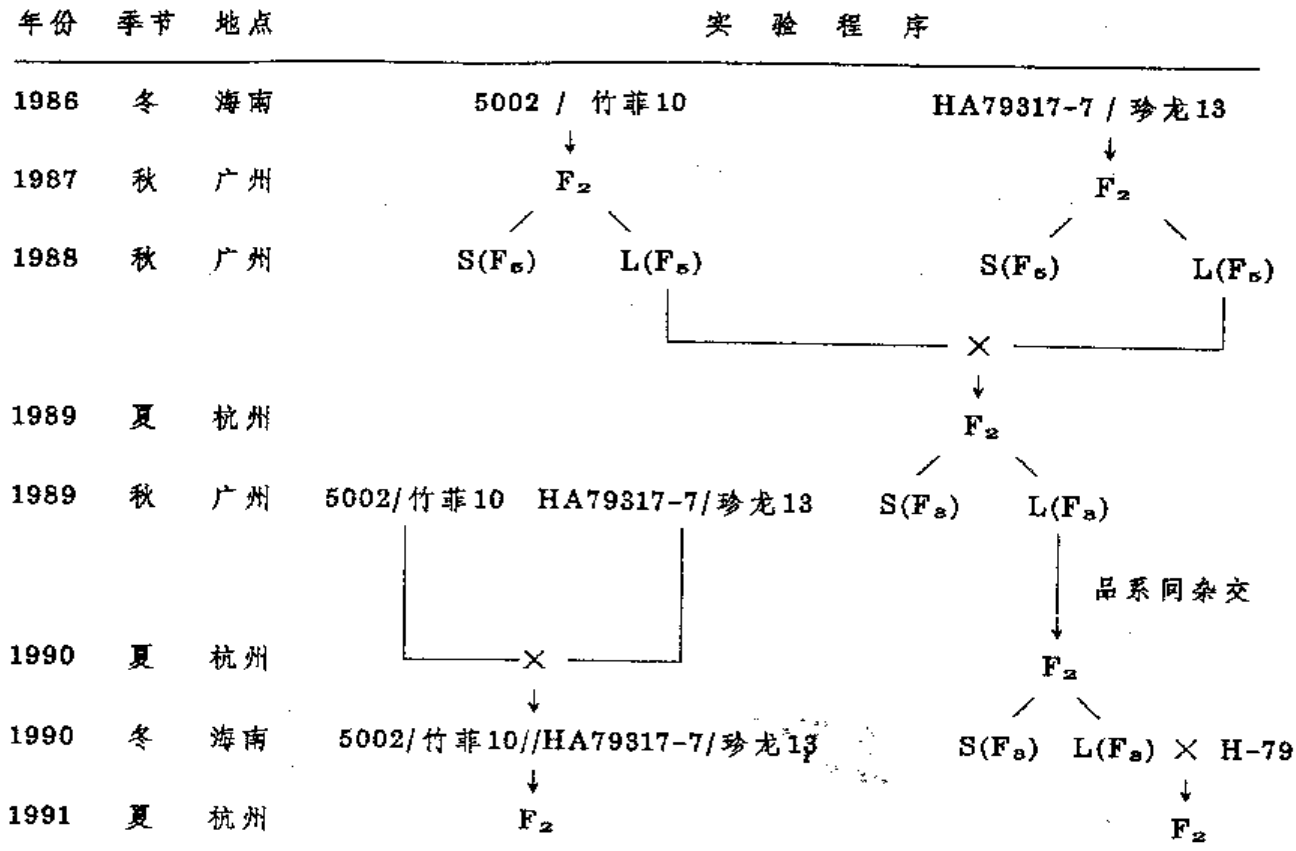


图11.5 从基因分散态组合中鉴别单个QTL

四个水稻品种的三轮杂交和歧化选择：L—大分蘖角品系；S—小分蘖角品系；H-79—由单基因 *la* 控制的匍匐型突变体

与上一轮所获相应的极端品系相似。由此推论，第二轮所得两类极端品系可能分别集中了4个原始亲本中所有QTL的增效和减效等位基因。利用第二轮所选极端品系进行相互杂交，并对其 F_2 群体进行第三轮歧化选择，未能获得超过亲本品系的遗传变异。经第二轮和第三轮选择所获的大分蘖角度品系，其分蘖角度最大达56.5°，这与受单基因 *la* 控制的匍匐型突变体 H-79的表型已十分接近。因此，通过三轮的歧化选择，在4个原始亲本间发现了影响分蘖角度的4个 QTL，而且对原始亲本及其分裂选择品系的6个群体进行遗传模型检验证实了这些 QTL 之间只存在加性效应，而与显性和上位性效应无关。将带有4个QTL全部增效等位基因的极端品系与突变体 H-79杂交，其 F_2 群体分蘖角的变异表现出主基因和数量性状基因所共同决定的特征 (Xu和Shen 1992c，表11.1)。

为探讨将分散分布于不同遗传材料中的多个QTL 同效等位基因累加于同一遗传品系之中的有效方法，我们以上述4个原始品种为亲本，先配制 5002/竹菲10和 HA79317-7/珍龙13两个单交组合，然后采用这两个组合进行复合杂交，从复交 F_1 中选择分蘖角度较大的 10个单株，当选个体自交产生复交 F_2 家系，从每一家系选择分蘖角度最大的一个个体。结果表明，最终当选的10个个体，其平均分蘖角度为 54.8°，与通过三轮歧化选择所得极端品系的分蘖角度相近。但复合杂交选择所需世代数由歧化选择的8～9代减少到5～6代。显

表11.1　不同遗传材料及其杂种后代中分蘖角的表现

群　体	样本数	均值	方差	分蘖角分布																	
				0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
HA79317-7 (P_1)	33	16.4	4.31			5	24	1													
竹菲10 (P_2)	24	12.9	2.77			21	3														
珍龙 13 (P_3)	29	12.1	2.93		4	24	1														
5002 (P_4)	30	15.5	4.17		4	24	1														
EG10	33	4.7	0.63	18	25																
LG19-1	39	55.9	7.92											15	19	4					
EG10/LG19-1 F_1	31	28.7	5.88						25	6											
EG10/LG19-1 F_2	186	32.2	109.87	1	2	4	14	25	35	34	29	16	15	7	3						
$P_2/P_4//P_3/P_2$	207	19.8	35.60	2	11	25	65	64	35	5											
家系合并群体	300	30.3	117.32	1	6	12	28	42	76	52	25	21	21	9	4	1					
H-79	22	89.0	1.99																		22
LG19-1/H-79 F_1	37	34.7	9.95						4	12	13	2									
LG19-1/H-79 F_2	215	43.9	495.85	1		2	5	15	25	45	44	14	12	5	4	1	2			2	34

EG10和LG19-1分别为第三轮歧化选择所得最小分蘖角和最大分蘖角品系；家系合并群体由$P_2/P_4//P_3/P_2F_1$中分蘖角度较大的10个个体的自交后代(每家系30株)混合而成。

然在多基因之间不存在连锁的条件下，采用复合杂交可以在较短的时间内将不同材料中的同效多基因累加在一起。但当连锁存在时，歧化选择将有助于打破连锁，促进基因的重组。

实践证明，在水稻等作物的不同品种间，普遍存在QTL同效等位基因的分散分布。例如以两个早熟籼稻品种 H1459和乐一为测验系，分别与泸红早等22个早籼品种进行简化三重测交分析，发现倒三叶长和剑叶夹角在两个测验系间可能存在由于同效等位基因分散分布引起的遗传差异(徐云碧和申宗坦1992b)。对水稻产量构成性状的分析表明，这两个测验系之间还可能存在控制单株草重的同效基因的分散分布(徐云碧和申宗坦 1991)。对水稻分蘖角度的另一研究表明，分蘖松散型品种矮脚南特和二九南1 号之间在两个基因座位上存在同效等位基因的分散分布，两个品种各带有三个基因座位的增效等位基因 (徐云碧和申宗坦 1993)。

以上研究也为我们今后分解多个QTL 提供了另一条思路。 如果能找到数量性状表现为两种极端类型的遗传材料，就可以将其相互杂交产生F_1，然后在F_2群体中通过选择获得具有不同QTL 等位基因的分离体。利用分离体相互杂交，根据 F_2代目标性状的分离即可鉴别QTL 等位基因的异同，从而将不同的QTL 分解开来，获得 QTL 克隆所需、带有不同QTL等位基因的遗传材料。

11.5 QTL克隆的可能途径

11.5.1 以图谱为基础的 QTL 克隆

对于具有类似于主基因效应的 QTL，其基因克隆可以完全采用类似于质量基因的克隆步骤进行。对于效应较小的 QTL，要尽量地控制外界环境对 QTL 效应的干扰，可采取在不同环境重复评价或在同一环境下设置有重复的随机区组试验来控制环境效应，提高性状评价的精确性，从而实现 QTL 的精确定位和有效地验证转数量基因植株的QTL 效应。采用只有目标数量性状存在差异的成对（近）等基因系分别用作有利基因的供体和克隆数量基因的供体，可能是提高效应较小的 QTL 的克隆效率的有效手段之一。

11.5.2 以随机 cDNA 序列分析为基础的 QTL 克隆

在11.3节中介绍的随机 cDNA 序列，是建立在表达基因的 cDNA 文库基础上的。在尝试用这一途径克隆QTL 时，首先应将已测序的 cDNA 尽可能地用作遗传图谱的 RFLP 标记或 PCR 标记，并用cDNA 标记饱和的连锁图谱定位各种数量性状的基因，这样就有可能在精确定位的基础上找到可能与某一种数量基因效应直接有关的 cDNA；再进一步通过转化，就可能证明该 cDNA 的基因是否为那种数量性状基因。此外，随着反向遗传学技术的发展，也可能通过 cDNA 序列的定点突变和基因工程的手段，研究突变的 cDNA 的遗传效应及其对数量性状的影响，进而确定数量性状的 cDNA。

广义的 cDNA 序列也包括在真核生物基因组研究中已积累的贮存在 DNA 数据库中的大量可编码序列，对这些序列的功能研究必将导致大量 QTL的鉴别和认定。如酵母第3染色体的315kb的全序列已经全部测定，约含 182个可编码 100 个氨基酸以上的可能的基因，可以推测，在这些基因中也包括了控制数量性状的多基因。

11.5.3 利用插入突变克隆 QTL

可以利用克隆的插入因子作为分子探针来鉴别影响数量性状的基因座位，以及为分离有关基因进行基因组文库的筛选。也有可能利用转座子移动或利用适当的载体将DNA 序列直接插入在基因组中的新位点。反过来，这又会导致产生影响数量性状的突变事件。如果插入因子已被克隆，并且影响目标数量性状的插入事件已被鉴定，就可以利用克隆的插入因子来拯救和克隆受影响的基因座位。当然，成功地应用这些技术取决于对插入突变造成的数量性状效应的精确判断和遗传分析。

利用插入突变克隆特定基因座位有些类似于“大海捞针”。也就是说，基因组中存在数量极大的插入位点，但能影响特定基因功能的位点数目有限。对于2×10^9碱基对的典型植物基因组，假定有效基因大小为$4\times10^3\sim4\times10^4$bp，且影响某个数量性状的基因座位数为10～100 。 为了有较大的概率包括影响一个数量性状的一个插入事件， 就必须检测约1000～10000个插入事件对十个数量性状的影响。要降低每个受影响QTL 所要检测的插入事件数，一种方式是利用基因组较小的植物，如拟南芥，其二倍体基因组为7×10^7bp，是一个特别有吸引力的植物。一旦在拟南芥或某一其它实验植物中克隆到QTL， 就可以根据

不同物种同类基因的同源性，通过分子杂交克隆不同物种的相应基因座位。

利用插入突变克隆QTL的第二个主要问题是突变等位基因效应的检测。如上所述，任一特定QTL的等位基因效应可能较小，尽管插入所产生的突变比自然群体经自然选择生存下来的QTL等位基因有更大的效应，但其效应仍不能与其它分离基因或环境变异的效应区分开来。其它基因座位的遗传变异可以通过自花授粉或高度近交品系内的插入突变来控制。多次重复鉴定每一插入因子可以控制剩余的环境变异。每株产生多重插入可减少所需鉴定的植株总数，这样每一重复单株可以同时鉴定多个插入因子。在此过程中分子标记方法可能是必需的，因为借助于分子标记可在鉴定和克隆计划中以及在后续QTL克隆中跟踪每一插入因子。因此在原理上，涉及克隆DNA的插入突变，首先可作为插入因子，然后作为分子标记探针来检测插入的存在，从而使数量性状基因的克隆成为可能。

11.5.4 其它方法

Robertson (1985) 提出了另一种可能用于分离数量基因的方法。这一方法是建立在下列假设基础上的，即影响数量性状的质量突变基因（如矮生、窄叶等突变体）代表了等位基因谱的一个极端，而另一极端是野生型等位基因。由于这些性状的质量突变基因可能同时影响着同类的数量性状，因而很明显这些座位上的野生型等位基因就负责着数量性状的野生型表型。Robertson (1989) 给出了两个支持上述假设的例子。其中之一是玉米中由主基因控制的GA产物缺乏程度的差异导致突变体株高的数量差异。他还列举列了一系列与数量性状表现相关联的同类质量性状（表11.2）。Beavis等 (1991) 发现所鉴别的有关玉米株高的大多数 QTL位于已定位的质量基因座位附近。Edwards等 (1992) 也发现许多影响玉米株高的 QTL分别定位在已知的影响株高的主基因座位上，表明有些QTL与影响株高的主基因可能是等位的。这种主微基因同处“一地”或“同源”现象已得到越来越多的实验结果的支持。在水稻稻瘟病抗性的遗传分析方面，朱立煌等 (1994) 对DH群体中抗病和感病株系进行 RAPD 分析将一个抗病的主基因定位在第8染色体上。Wang等 (1994) 在定位表现数量抗性的稻瘟病抗性基因时，在第5染色体上的同一区域也发现了与抗性有关的QTL。Helentjaris等 (1992) 认为 Robertson (1985) 的理论为重要数量性状的基因鉴别、克隆及遗传工程操纵提供了可能的途径，并以玉米的株高数量性状为模式，着手对有关问题进行研究。我们也许可以期望将过去已鉴别的基因与 QTL联系起来或将 QTL等价于传统图谱上的类似极端突变体，然后证实这种关系，并通过相关质量基因的克隆实现对QTL的克隆。

考虑到植物和动物中不同生理特性间的相关性，采用上述 QTL 克隆方法从影响这些生理特性的多基因系统中克隆一个或少数几个与其中某一生理特性有关的基因座位就可能对该多基因系统实现很大程度的遗传操纵。也就是说，迫使与多个生理特性相关的多基因系统的某一功能进入一个高速运转的“传动装置”，可能会带动该系统的其余功能与其一起运转。在生长激素转化小鼠中观察到这一现象：生长激素尽管只是发育的成分之一，但当通过插入额外拷贝的生长激素基因迫使其大量产生之后，生长激素足以使所有其它系统作出适当的反应。因此，只克隆影响某一数量性状的少数几个多基因座位，就可能足以使其它有关的基因座位作出连锁反应。而且，基于在不同物种同源基因之间存在高度的序列同源性，在一个物种中克隆的QTL就可能用于鉴定、定位和克隆其它近缘物种的同源基

表 11.2　玉米中可能与数量性状相关联的同类质量性状

质量性状	数量性状	不同的质量突变体数目
类矮秆 (*d*, *an*, *py*, *br*)	株高	17+
叶绿素缺乏 (*w*,*l*,*v*,*pg*)	叶绿素含量	30+
强叶绿素荧光 (*hcf*)	光合成效率	30+
类胡萝卜素缺乏 (*lw1*,*vp2*,*w3*)	植株和籽粒中的类胡萝卜素含量	15+
无胚籽粒 (*gm*,*mg*)	籽粒油分含量	未研究但观察到大量
籽粒缺陷 (*dek*,*cp*,*mn*)	籽粒大小	30+
皱缩籽粒 (*sh*,*su*,*bt*,*ae*,*wx*)	淀粉产生	12+
无花(空秆) (*ba*,*bs*)	穗大小	3+（可能很多）
光泽叶 (*gl*)	蜡质产生	20+
不限定生长 (*id*,*ga1*,*gt*)	成熟时间	3+（可能更多）
窄叶 (*nl*)	叶面积	未知，但可能很多

因座位。似乎有理由期望，在许多情况下具有同源核苷酸序列的基因座位将具有类似的功能，所以在一个物种中克隆的QTL也可为近缘物种同源QTL的克隆提供非常直接的鉴定。

11.6　与QTL克隆有关的基础研究

从整体上看，至少在目前，上述有关的QTL克隆策略还存在许多尚待解决的问题。首先，数量性状的遗传表达一般涉及一个以上的基因座位。在理论上，我们可以采用一般的基因克隆方法将效应较大的QTL一一克隆出来，但如何将这些克隆基因组装成一个整体，使之在新的遗传背景下实现正常的表达则是一个尚待探索的难题，否则就难以证明多个QTL的整体效应。

其次，许多所谓的数量性状实质上是许多不同或相关性状构成的复合体，如作物的产量性状是由各个产量构成因素（性状）综合而成的，而种子油分或蛋白质的含量则与多种化合物和氨基酸有关。因此，数量性状本身的精细分解和分析，是这类数量性状基因克隆的基础。这种分解和分析的方法涉及生化、生理、解剖、细胞、分化、发育等许多不

同学科的研究手段。 将同一数量性状剖分为几个不同的组分，也有利于基因的定位与克隆。 如将作物由多基因造成的低结实性剖分为由于雄性不育、雌性不育引起的低结实，或由于胚囊败育、花粉败育等不同原因引起的低结实，就可能将多个遗传因子分解为起不同作用的基因。

第三，数量性状的表达同时受外界环境和生物体内部遗传背景的较大影响。外部环境（包括水分、光照、温度、湿度、各种土壤因子等）对数量性状的表现型起着重要的修饰作用。在很多情况下，外界环境成为性状表达的“调节因子”，如在水稻中发现的各种光（温）敏不育系，其不育性的表现就受外界光（温）条件的制约。 来自同一对亲本的遗传群体，在不同的环境条件下对数量性状进行评价，采用前述的QTL 作图方法进行作图，发现所检测到的 QTL 只有一部分是可在不同环境下同时检测到的，而其它的则与特定的环境相联系，表现出环境依赖性。因此，为了有效地进行 QTL 的作图和克隆，还必须研究不同数量性状表达所需的特定环境条件，将综合的环境因子分解为具体而又精确的外界条件；并使不同遗传材料间目标数量性状的差异在受控的特定条件下达到最大，即最大限度地区分出 QTL 的效应。

第四，数量性状基因的表达在很大程度上受到遗传背景的影响。这表现在控制不同的数量性状的基因之间，数量基因与其它主效基因之间，它们的作用不是相互独立的。因此，当同一数量基因处于不同的遗传背景时，其表达的程度是有差异的。遗传背景的异质性是造成不同遗传实验结果难以相互验证的重要原因之一。在质量性状的遗传研究中，采用只具有目标性状存在差异的（近）等基因系（参见第5章）作为遗传研究的材料，可以大大降低遗传背景的异质性，从而更精确地研究目标基因的效应。与此相类似，在分子数量遗传研究中，可以通过培育遗传背景基本一致（同质）的近等数量基因系以减少遗传背景的干扰。不过，数量性状近等基因系的培育是一项费时而且效率较低的工作，目前尚未见有关的报道。

第五，与质量性状类似，数量性状的遗传表达是与发育相联系的，存在表达的发育阶段性。研究数量性状表达的发育遗传学问题，不仅是经典的数量遗传分析所必需的，而且有助于确定数量性状遗传差异表现的最佳时期，从而为 QTL 作图和克隆奠定基础。我们通过对水稻不同发育阶段分蘖数的遗传分析，证实了同一数量性状在不同发育时期表达的差异，确定了对分蘖数进行评价和选择的适当时期为分蘖盛期 (Xu和Shen1991)。 在这一时期，不同遗传材料之间分蘖数的差异达到最大，是区分各种可能的 QTL基因型的最佳时期。 可以说，对目标数量性状进行发育遗传学的深入研究是 QTL 克隆的基础性工作，而这些工作目前还做得很少。

此外，根据 Robertson (1985) 提出的设想进行数量基因的分离，还必须对影响同一数量性状的主效基因和微效基因进行研究，以探明它们对性状表达的相对贡献、两者间的相互关系和相互作用。在统计分析上，还需发展与之相适应的、可以鉴别不同类型基因、可估计基因效应的遗传模型和方法，以及发展借助于分子标记进行主-微效应QTL作图的方法。 目前这类问题及其相关研究已引起分子遗传学和数量遗传学两个领域科学家的重视 (莫惠栋1993a,b; Eltson 1984; Jiang 等1994)。

在理论上，我们可将分子生物学领域发展的各种分子克隆技术应用于数量性状的遗传

操纵。但在实践上，由于数量性状的评价对于群体的依赖以及外界环境对数量表型的干扰，因而需要发展与之相适应的研究方法。

附　　录

附录部分包括三方面的内容：QTL 作图软件，进一步阅读的参考文献，以及练习题。我们希望通过这个附录为读者提供更多的信息。QTL 作图软件作为我们近几年的研究成果奉献给读者。我们收集了反映分子数量遗传学相关领域及最新进展的参考文献，以期为读者提供更广泛的背景材料。为了帮助初学者通过实践来巩固有关的知识，我们还编写了若干练习题供参考。

A1　QTL 作图软件

附录 A1 提供了本书第6～8章所介绍的常用QTL 作图方法的计算机程序，包括用于QTL 作图的数据文件的建立，采用标记基因型之间数量性状的平均值比较法、单因素方差分析法、以单一标记为基础的矩估计法和最大似然法以及以两个相邻标记为基础的区间作图法进行 QTL 作图。全部程序采用Quick BASIC 语言设计，可在 PC 286及其以上的微机上应用。其中 DATASET.BAS 为后续所有程序建立数据文件，其中的DATA 来自我们的研究资料，与第6～8章所引实例完全相同。读者若要分析自己的数据，只需替换其中的DATA 语句。所要分析的数据可来自某个具体的实验，也可以是通过计算机模拟所产生的数据。为了减少程序录入的错误，读者可来信索取录有这些程序的软盘。

A1.1　建立数据文件 (DATASET.BAS)

程序名称：DATASET.BAS

程序用途：为后续的其它程序建立可以调用的顺序文件，并同时输出整理好的数据。程序运行的结果之一是建立两个顺序文件，一个名为 "Traitdat"，用于存放数量性状观测值，一个名为 "Markdat"，用于存放分子标记数据。

DATA 语句的说明：

用于 QTL 作图的基本数据包括作图群体的数量性状表型值和分子标记数据。数据以群体中的分离单位为基础，例如 F_2 和BC群体以单株为单位，而 DH 群体以株系为单位。如果 DH 群体在具体观测数量性状时每个株系考查了多个植株，则以多个植株观测值的平均值来替代F_2群体的个体表型值。几个赋值语句说明如下：

85　作图群体的种类，可为 F_2、BC或 DH。

86　产生作图群体的两个亲本的名称，依次为母本、父本

87　标记基因型种类的代码，第一个数字代表母本基因型，第二个数字代表父本基因型，第三个数字代表杂合体类型，第四个数字代表因各种可能的原因所造成的数据缺失。一般这四个数字依次为 1,3,2,0。

90～100　数量性状的名称

1011～ 1071 数量性状表型值，排列顺序为第一植株或株系的第1，2，…， c 个性状，然后是第二植株的第1，2，…， c个性状，… 。

1110～1170 分离单位的标记基因型种类，其中第一行的数据依次为染色体编号、标记或用于分子杂交的探针名称、与探针相结合的限制性内切酶的名称。随后的数据为群体中各个分离单位的标记基因型。

程序：

```
10 ' Program for transformation of trait and marker data into files,
   ' Traitdat and Markdat, which will be used by other programs

15 INPUT "Population size; n= "; n
   INPUT "Number of quantitative trait; c= "; c
20 INPUT "Number of molecular markers; m= "; m

25 DIM trait$(c), trait(n, c), chr(m), pro$(m), RFLP(m, n)
   p5$ = " ### ### ##.# ### ###.##"
   p$ = "------------------------------------------------------------"

30 OPEN "Traitdat" FOR OUTPUT AS #3

   READ Pop$, p1$, p2$, hompl, homp2, hetero, missing
   WRITE #3, Pop$, p1$, p2$, hompl, homp2, hetero, missing, m, n, c
   FOR i = 1 TO c: READ trait$(i): WRITE #3, trait$(i): NEXT i
   FOR i = 1 TO n
   FOR j = 1 TO c
   READ trait(i, j): WRITE #3, trait(i, j)
   NEXT j, i
40 CLOSE

50 OPEN "Markdat" FOR OUTPUT AS #3
   FOR i = 1 TO m
   READ chr(i), pro$(i), en$ : WRITE #3, chr(i), pro$(i)
   FOR j = 1 TO n
   READ RFLP(i, j):WRITE #3, RFLP(i, j)
   NEXT j, i
60 CLOSE

70 OPEN "rawdata" FOR OUTPUT AS #4
   PRINT :PRINT :PRINT #4, :PRINT #4,
   PRINT #4, "Phenotypic values of quantitative traits in the mapping population"
```

```
PRINT #4, p $
PRINT "Phenotypic values of quantitative traits  in the mapping population"
PRINT p $

FOR i = 1 TO n
PRINT #4, USING "## "; i;
PRINT #4, USING p5 $ ; trait(i, 1);  trait(i,   2);  trait(i,   3);  trait(i,   4);  trait(i,   5)
PRINT USING "## "; i;
PRINT USING p5 $ ; trait(i, 1);  trait(i,   2);  trait(i,   3);  trait(i,   4);  trait(i,   5)
NEXT i
PRINT #4, p $ : PRINT #4, "Trait no:"
PRINT p $ : PRINT "Trait no:"
FOR i = 1 TO c
PRINT #4, i; TAB(5);"---";  :  PRINT #4, trait $ (i)
PRINT i; TAB(5);"---";  :  PRINT trait $ (i)
NEXT i

PRINT #4, : PRINT #4, : PRINT #4, : PRINT #4,
PRINT : PRINT : PRINT : PRINT

PRINT #4, "Molecular genotype of plants or lines"
PRINT #4, "      in the mapping population"
PRINT #4, p $
PRINT "Molecular genotype of plants or lines"
PRINT "      in the mapping pupolation"
PRINT p $
FOR i = 1 TO n
PRINT #4, USING "##"; i;
PRINT #4, TAB(4);
PRINT USING "##"; i;
PRINT TAB(4);
FOR j = 1 TO m
PRINT #4, USING "####"; RFLP(j, i);
PRINT USING "####"; RFLP(j, i);
NEXT j: PRINT : PRINT #4,
NEXT i
PRINT #4, p $
PRINT #4, "Mapping population: "; p1 $ ; "/"; p2 $ ; " "; Pop $
PRINT #4, "Population size; n="; n
PRINT #4, "Number of quantitative trait; c="; c
PRINT #4, "Number of molecular markers; m="; m
```

```
    PRINT #4, "Marker genotype: "; hompl; ": "; p1 $
    PRINT #4, "                 "; homp2;": "; p2 $
    PRINT #4, "                 "; hetero; ": "; "F1"
    PRINT #4, "                 "; missing; ": "; "Missing"
    PRINT p $
    PRINT "Mapping population: "; p1 $ ;"/"; p2 $ ; " "; Pop $
    PRINT "Population size; n="; n
    PRINT "Number of quantitative trait; c="; c
    PRINT "Number of molecular markers; m="; m
    PRINT "Marker genotype: "; hompl; ": "; p1 $
    PRINT "                 "; homp2;": "; p2 $
    PRINT "                 "; hetero; ": "; "F1"
    PRINT "                 "; missing; ": "; "Missing"

 80 END

 85  DATA "F2"
 86  DATA "Zhai-Ye-Qing 8", "Jing-Xi 17"
 87  DATA 1,3,2,0
 90  DATA "Days to heading"
     DATA "Plant height (cm)"
     DATA "Length of panicle (cm)"
     DATA "Number of spikelets/panicle"
100  DATA "Fertility(%)"

1001  DATA 126,115,20.5,129,42.64,107,127,27.5,356,44.94
      DATA 112,123,29.0,341,25.81,95,117,20.5,202,60.89
      DATA 115,121,24.0,265,26.67,110,127,24.3,311,34.73
      DATA 101,103,20.7,126,42.06,119,116,25.2,237,46.84
      DATA 118,112,22.0,217,56.68,115,111,24.7,186,55.91

1011  DATA 104,115,27.5,330,28.79,109,106,22.5,162,59.88
      DATA 107,115,26.0,270,40.00,114,123,23.7,314,44.27
      DATA 123,97,20.7,146,49.32,111,111,24.9,254,35.04
      DATA 119,110,22.0,322,72.05,113,126,28.8,357,28.57
      DATA 124,102,25.7,223,18.39,102,110,23.0,231,45.02

1021  DATA 116,126,24.8,310,21.94,113,111,19.5,201,47.26
      DATA 115,132,28.3,411,27.49,101,102,22.8,171,32.16
      DATA 109,120,23.4,233,36.05,111,138,26.5,294,61.90
      DATA 110,122,29.9,224,60.27,115,123,27.5,233,47.21
```

```
      DATA 114,111,24.0,197,28.93,109,119,26.5,237,39.66

1031  DATA 114,119,24.5,221,49.77,119,109,20.0,273,37.00
      DATA 105,126,30.6,245,28.98,113,134,27.0,327,55.05
      DATA 119,110,25.7,360,80.56,116,123,24.0,217,71.43
      DATA 107,118,23.4,183,64.48,114,123,25.5,207,61.69
      DATA 124,107,21.7,224,69.64,109,106,25.4,254,60.63

1041  DATA 106,123,25.2,186,60.22,106,108,20.4,115,78.26
      DATA 92,100,24.2,260,82.69,115,112,20.5,188,52.66
      DATA 113,113,24.0,270,61.11,108,113,25.5,242,15.70
      DATA 125,106,21.8,158,44.94,120,105,19.5,126,38.68
      DATA 101,110,22.5,215,34.72,121,104,23.5,161,79.50

1051  DATA 124,91,21.5,189,16.93,122,110,21.5,182,43.96
      DATA 112,111,21.5,186,77.42,114,127,23.1,180,84.44
      DATA 106,97,18.1,162,27.16,124,122,24.2,214,72.90
      DATA 89,108,23.0,172,47.67,112,110,23.2,198,56.57
      DATA 114,140,27.0,202,73.27,118,94,20.0,187,86.10

1061  DATA 106,110,22.7,176,81.82,117,106,25.0,227,36.56
      DATA 114,96,22.2,154,13.64,111,107,25.5,220,76.35
      DATA 119,116,24.5,189,43.92,109,90,21.5,145,23.45
      DATA 108,95,22.1,149,64.43,104,120,22.5,167,52.10
      DATA 103,101,20.5,134,76.12,99,104,23.5,225,80.00

1071  DATA 109,124,25.5,210,43.10,115,118,28.7,237,59.49
      DATA 112,118,22.3,198,30.30,111,135,28.9,213,49.77
      DATA 114,100,19.8,155,6.45,117,131,26.5,327,44.34

1110  DATA 5,RG573,HindIII
      DATA 1,2,2,1,3,2,3,2,1,3,2,2,2,3,1,3,2,1,2,2,2,2,2,3,1,2,1,2,2,2
      DATA 2,2,2,2,3,3,3,2,3,2,2,1,2,1,2,2,2,3,2,2,2,3,3,3,3,1,3,2,1,2
      DATA 1,3,3,3,1,3,1,2,3,2,2,1,3,2,1,1

1120  DATA 5,RG13,HindIII
      DATA 1,2,1,1,3,2,3,2,2,3,3,2,2,3,1,3,3,0,1,2,2,2,3,2,1,1,1,2,3,2
      DATA 2,3,2,2,3,2,2,2,3,2,3,3,2,3,2,2,1,3,2,2,2,3,3,3,3,1,3,2,1,2
      DATA 1,2,2,3,1,3,1,2,3,2,2,2,3,2,1,2

1130  DATA 5,RZ70,HindIII
      DATA 3,2,1,3,1,1,2,2,3,3,3,3,2,2,3,2,2,2,1,1,2,2,2,2,2,2,2,2,2,3
      DATA 2,2,2,2,2,2,3,2,3,2,2,3,3,3,2,2,2,2,2,3,2,3,3,3,3,2,3,2,2,2
      DATA 1,3,3,3,1,3,2,2,3,2,3,1,3,2,2,1
```

```
1140 DATA 5,RG697,EcoRV
     DATA 1,3,2,1,3,2,2,0,0,0,0,2,0,3,0,3,2,2,3,0,1,2,2,1,1,0,1,2,2,3
     DATA 2,2,3,2,2,2,2,2,2,2,2,2,2,2,1,2,1,2,2,2,2,2,2,3,2,3,2,2,3
     DATA 2,3,3,3,1,3,1,2,2,3,3,0,3,1,3,1

1150 DATA 5,RG474,XbaI
     DATA 1,3,1,2,2,1,1,3,2,3,2,2,2,3,1,2,2,2,2,2,2,3,2,3,0,2,1,2,2,3
     DATA 1,2,3,2,2,2,2,2,2,2,2,3,2,3,2,2,2,1,2,1,2,3,2,1,3,2,3,2,2,3
     DATA 2,3,2,3,1,3,1,2,2,3,3,1,2,3,3,1

1160 DATA 5,RG435,HindIII
     DATA 1,3,1,1,3,2,2,3,2,3,2,2,2,3,1,3,2,0,3,2,2,3,3,3,1,1,2,2,2,3
     DATA 2,2,3,2,2,2,1,2,1,2,2,3,1,3,2,2,2,1,2,1,2,3,2,1,3,1,3,2,2,3
     DATA 2,3,2,3,1,3,1,2,2,3,3,1,2,2,3,1

1170 DATA 5,RZ225,XbaI
     DATA 1,3,2,1,3,2,2,3,3,3,2,2,2,3,2,3,2,1,3,2,1,3,0,3,1,2,1,2,2,3
     DATA 2,2,2,2,2,2,2,2,3,2,2,1,1,1,2,1,2,1,2,2,2,3,2,2,3,2,3,1,2,3
     DATA 2,3,3,3,1,3,1,2,2,3,3,1,2,2,3,1
```

程序运行时，首先要求输入群体大小（n）、待分析的数量性状的数目（c）和分子标记数(m)。运行结束时将产生"Traitdat"和"Markdat"两个顺序文件，同时输出以下结果（这些结果同时存入名为"rawdata"的文件，可供DOS的EDIT命令或其它编辑软件所调用。

运行结果如（n=76，c=5，m=7）：

Phenotypic values of quantitative traits in the mapping population

1	126	115	20.5	129	42.64
2	107	127	27.5	356	44.94
3	112	123	29.0	341	25.81
4	95	117	20.5	202	60.89
5	115	121	24.0	285	26.67
6	110	127	24.3	311	34.73
7	101	103	20.7	126	42.06
8	119	116	25.2	237	46.84
9	118	112	22.0	217	56.68
10	115	111	24.7	186	55.91
	...	...	...		

75	114	100	19.8	155	6.45
76	117	131	26.5	327	44.34

Trait no:

1 ---Days to heading
2 ---Plant height (cm)
3 ---Length of panicle (cm)
4 ---Number of spikelets/panicle
5 ---Fertility(%)

Molecular genotype of plants or lines
in the mapping pupolation

1	1	1	3	1	1	1	1
2	2	2	2	3	3	3	3
3	2	1	1	2	1	1	2
4	1	1	3	1	2	1	1
5	3	3	1	3	2	3	3
6	2	2	1	2	1	2	2
7	3	3	2	2	1	2	2
8	2	2	2	0	3	3	3
9	1	2	3	0	2	2	3
10	3	3	3	0	3	3	3
	...		...		...		
75	1	1	2	3	3	3	3
76	1	2	1	1	1	1	1

Mapping population: Zhai-Ye-Qing 8/Jing-Xi 17 F2
Population size; n= 76
Number of quantitative trait; c= 5
Number of molecular markers; m= 7
Marker genotype: 1 : Zhai-Ye-Qing 8
3 : Jing-Xi 17
2 : F1
0 : Missing

A1.2 均值比较 (QTLMEAN.BAS)

程序名称: QTLMEAN.BAS

程序用途: 通过t测验，比较标记基因型之间数量性状的平均值，进行标记-数量性状之间相互关联的检测。

DATA 语句: 所列数据为t测验所需的显著性值。

程序:

```
10  ' Program for detection of phenotypic difference between alternative
    ' marker genotypes by t-test
    ' This program uses data files  set by the program DATASET.BAS

    WIDTH "LPT1:", 250

30  OPEN "Traitdat" FOR INPUT AS #3
    OPEN "qtlmean" FOR OUTPUT AS # 4

    INPUT #3, pop $ , p1 $ , p2 $ , homp1, homp2, hetero, missing
    INPUT #3, m, n, c
    IF pop $ = "F2" THEN ww = 3 ELSE ww = 2

40  DIM trait $ (c), trait(n, c), ch $ (c), pro $ (m)
    DIM mark(m, n), chr(m), m3(3, c), n(3), x(3), s(3)
    DIM xxx(3, n), t4(3, c), df4(3, c), t(20), t1(200), t2(200), tt(c)

50  FOR i = 1 TO 30 : READ t1(i), t2(i) : NEXT i
    READ t1(40), t2(40), t1(60), t2(60), t1(120), t2(120), t1(200), t2(200)

60  bb $ = "------------------------------------------------------------------------"
    PRINT #4,"Detection of phenotypic difference between alternative marker genotypes"
    PRINT #4, " by t-test (** and * show significance at 1% and 5% levels,"
    PRINT #4, " respectively; trait numbers are 1, 2, ...) "
    PRINT #4, bb $
    PRINT #4, "Chr Marker ";
    PRINT " Detection of phenotypic difference between alternative marker genotypes"
    PRINT " by t-test (** and * show significance at 1% and 5% levels,"
    PRINT " respectively; trait numbers are 1, 2, ...) "
    PRINT bb $ : PRINT "Chr Marker ";
    FOR i = 1 TO c
    PRINT #4, TAB(13 + i* 9); i; : PRINT TAB(13 + i * 9); i;
    NEXT i
```

```
    PRINT #4, : PRINT : PRINT #4, bb $ : PRINT bb $

70  FOR i = 1 TO c : INPUT #3, trait $ (i) : NEXT i

    FOR i = 1 TO n : FOR j = 1 TO c
    INPUT #3, trait(i, j)
    NEXT j, i
    CLOSE #3

80  OPEN "markdat" FOR INPUT AS #3
    FOR i = 1 TO m : INPUT #3, chr(i), pro $ (i)
    FOR j = 1 TO n : INPUT #3, mark(i, j)
    NEXT j, i
    CLOSE #3

90  FOR ii = 1 TO m
    FOR jj = 1 TO c
    FOR j = 0 TO ww: n(j) = 0: x(j) = 0: s(j) = 0: NEXT j

    FOR i = 1 TO n
    IF mark(ii, i) > 3 THEN mark(ii, i) = 0
    n(mark(ii, i)) = n(mark(ii, i)) + 1
    x(mark(ii, i)) = x(mark(ii, i)) + trait(i, jj)
    xxx(mark(ii, i), n(mark(ii, i))) = trait(i, jj)
    s(mark(ii, i)) = s(mark(ii, i)) + trait(i, jj) ^ 2
    NEXT i
    x = 0: s = 0
    FOR i = 1 TO ww: x1(i) = 0: s1(i) = 0: NEXT i
    FOR i = 1 TO ww
    IF n(i) < 2 THEN 300
    NEXT i

    nn = n(1) + n(2) + n(3)
    FOR i = 1 TO ww : FOR j = 1 TO n(i)
    x = x + xxx(i, j) : s = s + xxx(i, j) ^ 2
    x1(i) = x1(i) + xxx(i, j) : s1(i) = s1(i) + xxx(i, j) ^ 2
    NEXT j, i
    sss = s - x ^ 2 / nn: sst = 0
    FOR i = 1 TO ww: sst = sst + x1(i) ^ 2 / n(i): NEXT i
    sst = sst - x ^ 2 / nn : sse = sss - sst
    FOR j = 1 TO ww
```

```
    s(j) = (s(j) - x(j) ^ 2 / n(j)) / (n(j) - 1) : x(j) = x(j) / n(j)
    NEXT j

    IF ww =2 THEN 95
    t4(2, jj) = ABS(x(1) - x(3)) / SQR(s(1) / n(1) + s(3) / n(3))
    s100 = (s(1) / n(1)) ^ 2 / (n(1) - 1) + (s(3) / n(3)) ^ 2 / (n(3) - 1)
    df4(2, jj) = (s(1) / n(1) + s(3) / n(3)) ^ 2 / s100
    t4(3, jj) = ABS(x(2) - x(3)) / SQR(s(3) / n(3) + s(2) / n(2))
    s200 = (s(3) / n(3)) ^ 2 / (n(3) - 1) + (s(2) / n(2)) ^ 2 / (n(2) - 1)
    df4(3, jj) = (s(3) / n(3) + s(2) / n(2)) ^ 2 / s200
95  t4(1, jj) = ABS(x(1) - x(2)) / SQR(s(1) / n(1) + s(2) / n(2))
    s300 = (s(1) / n(1)) ^ 2 / (n(1) - 1) + (s(2) / n(2)) ^ 2 / (n(2) - 1)
    df4(1, jj) = (s(1) / n(1) + s(2) / n(2)) ^ 2 / s300
    m3(1, jj) = x(1): m3(2, jj) = x(2): m3(3, jj) = x(3)
    tt(jj) = (t4(1, jj) + t4(2, jj) + t4(3, jj)) / 3
    NEXT jj

100 vv = chr(ii)
    IF vv <> w THEN PRINT USING "## "; chr(ii); ELSE PRINT " ";
    IF vv <> w THEN PRINT #4, USING "## "; chr(ii); ELSE PRINT #4, " ";
    PRINT #4, TAB(5); pro $ (ii); : PRINT #4, TAB(14); "1-2";
    PRINT #4, TAB(20); : PRINT TAB(5); pro $ (ii);
    PRINT TAB(14); "1-2"; : PRINT TAB(20);

    FOR j = 1 TO c
    PRINT #4, USING "###.##"; m3(1, j) - m3(2, j);
    PRINT USING "###.##"; m3(1, j) - m3(2, j);
    IF df4(1, j) > 60 THEN df4(1,j) = 120: GOTO 103
    IF df4(1, j) > 40 THEN df4(1,j) = 60: GOTO 103
    IF df4(1, j) > 30 THEN df4(1,j) = 40: GOTO 103
103 IF t4(1, j) > t2(df4(1, j)) THEN PRINT #4, "** "; : PRINT "** "; : GOTO 104
    IF t4(1, j) > t1(df4(1, j)) THEN PRINT #4, "* ";: PRINT "* "; : GOTO 104
    PRINT #4, "  "; : PRINT "  ";
104 NEXT j
    PRINT #4, : PRINT

    IF ww <> 3 THEN 200
    PRINT #4, TAB(14); "1-3"; : PRINT #4, TAB(20);
    PRINT TAB(14); "1-3"; : PRINT TAB(20);

    FOR j = 1 TO c
    PRINT #4, USING "###.##"; m3(1, j) - m3(3, j);
```

```
    PRINT USING "###.##"; m3(1, j) - m3(3, j);
    IF df4(2, j) > 60 THEN df4(2, j) = 120: GOTO 105
    IF df4(2, j) > 40 THEN df4(2, j) = 60: GOTO 105
    IF df4(2, j) > 30 THEN df4(2, j) = 40: GOTO 105
105 IF t4(2, j) > t2(df4(2, j)) THEN PRINT #4, "** "; : PRINT "** "; : GOTO 106
    IF t4(2, j) > t1(df4(2, j)) THEN PRINT #4, "* "; : PRINT "* "; : GOTO 106
    PRINT #4, "  "; : PRINT "  ";
106 NEXT j
    PRINT #4, : PRINT : PRINT #4, TAB(14); "2-3";
    PRINT #4, TAB(20); : PRINT TAB(14); "2-3"; : PRINT TAB(20)
    FOR j = 1 TO c
    PRINT #4, USING "###.##"; m3(2, j) - m3(3, j);
    PRINT USING "###.##"; m3(2, j) - m3(3, j);
    IF df4(3, j) > 60 THEN df4(3, j) = 120: GOTO 107
    IF df4(3, j) > 40 THEN df4(3, j) = 60: GOTO 107
    IF df4(3, j) > 30 THEN df4(3, j) = 40: GOTO 107
107 IF t4(3, j) > t2(df4(3, j)) THEN PRINT #4, "** "; : PRINT "** "; : GOTO 108
    IF t4(3, j) > t1(df4(3, j)) THEN PRINT #4, "* "; : PRINT "* "; : GOTO 108
    PRINT #4, "  "; : PRINT "  ";
108 NEXT j
    PRINT #4, : PRINT

200 w = vv
    PRINT #4, : PRINT

300 NEXT ii
    PRINT bb$ : PRINT #4, bb$

    CLOSE

    END

400  DATA 12.706,63.657,4.303,9.925,3.182,5.841,2.776,4.604,2.571,4.032
     DATA 2.447,3.707,2.365,3.499,2.306,3.355,2.262,3.250,2.228,3.169
     DATA 2.201,3.106,2.179,3.055,2.160,3.012,2.145,2.977,2.131,2.947
     DATA 2.120,2.921,2.110,2.898,2.101,2.878,2.093,2.861,2.086,2.845
     DATA 2.080,2.831,2.074,2.819,2.069,2.807,2.064,2.797,2.060,2.787
     DATA 2.056,2.779,2.052,2.771,2.048,2.763,2.045,2.756,2.042,2.750
     DATA 2.021,2.704,2.000,2.660,1.980,2.617,1.960,2.576
```

对于 F_2群体，有三种标记基因型，相互之间有三组比较。 其中 1-2，1-3和2-3分别

代表母本基因型的个体群与杂合体个体群之间，母本基因型个体群和父本基因型个体群之间以及杂合体个体群与父本基因型个体群之间数量性状均值的比较。运行结果给出相应的t值（此结果同时存入 "QTLMEAN" 文件供DOS 的EDIT 命令或其它编辑软件调用。

运行结果：

Detection of phenotypic difference betweem alternative markergenotypes by t-test (** and * show significance at 1% and 5% levels, respectively; trait numbers are 1, 2, ...)

Chr	Marker		1	2	3	4	5
5	RG573	1-2	2.21	-0.19	-1.06	-41.74*	6.55
		1-3	2.19	5.99	0.96	3.26	3.69
		2-3	-0.02	6.18*	2.03**	45.00**	-2.86
	RG13	1-2	3.84	0.24	-0.66	-28.32	0.51
		1-3	2.87	3.63	1.10	-16.71	-1.11
		2-3	-0.97	3.39	1.76**	11.62	-1.63
	RZ70	1-2	1.02	1.78	1.00	20.92	-5.73
		1-3	2.49	8.44*	2.75**	68.93**	-11.84
		2-3	1.47	6.67*	1.75**	48.01**	-6.11
	RG697	1-2	-1.04	1.85	0.46	-14.98	-13.67*
		1-3	1.50	6.75	0.23	-10.92	3.00
		2-3	2.54	4.90	-0.24	4.07	16.67*
	RG474	1-2	3.05	-0.79	-0.05	-31.95	1.45
		1-3	5.10*	4.63	0.87	-1.96	4.12
		2-3	2.05	5.42	0.92	29.98	2.67
	RG435	1-2	2.35	0.15	-0.40	-14.44	8.12
		1-3	2.98	3.97	-0.01	-9.70	13.84*
		2-3	0.63	3.82	0.40	4.75	5.73
	RZ225	1-2	-0.72	-1.40	-0.26	-7.13	-4.38
		1-3	-0.77	5.97	0.73	2.92	5.48
		2-3	-0.06	7.38*	0.99	10.04	9.86

A1.3 单因素方差分析 (QTLVAR.BAS)

程序名称：QTLVAR.BAS

程序用途：采用单因素方差分析方法，比较不同标记基因型之间数量性状的差异，以检测标记-数量性状的相互关联。

程序：

```
10  ' Program one-way analysis of variance between alternative  marker
    ' genotypes by F test
    ' This program  uses data files  set  by the program  DATASET.BAS

    WIDTH "LPT1:", 255
30  OPEN "qtlvar" FOR OUTPUT AS #4
    OPEN "Traitdat" FOR INPUT AS #3

    INPUT #3, pop $, p1 $, p2 $, homp1, homp2, hetero, missing
    INPUT #3, m, n, c
    IF pop $ = "F2" THEN ww = 3 ELSE ww = 2

40  DIM trait $ (c), trait(n, c), ch $ (c), pro $ (m)
    DIM mark(m, n), chr(m), m3(3, c), n(3), x(3), s(3), xxx(3, n), f(c)

60  bb $ = "------------------------------------------------------------------------"
    PRINT #4, " One-way analysis of variance among alternative marker genotypes"
    PRINT #4, bb $ : PRINT #4, "Chr Marker";
    PRINT " One-way analysis of variance among alternative marker genotypes"
    PRINT bb $ : PRINT "Chr Marker";
    FOR i = 1 TO c
    PRINT #4, TAB(16 + i * 9); i; : PRINT TAB(16 + i * 9); i;
    NEXT i: PRINT #4, : PRINT : PRINT #4, bb $ : PRINT bb $

70  FOR i = 1 TO c : INPUT #3, trait $ (i) : NEXT i
    FOR i = 1 TO n : FOR j = 1 TO c
    INPUT #3, trait(i, j)
    NEXT j, i
    CLOSE #3

80  OPEN "Markdat" FOR INPUT AS #3
    FOR i = 1 TO m : INPUT #3, chr(i), pro $ (i)
    FOR j = 1 TO n : INPUT #3, mark(i, j)
    NEXT j, i
```

```
    CLOSE #3

90  FOR ii = 1 TO m
    FOR jj = 1 TO c
    FOR j = 0 TO ww : n(j) = 0: x(j) = 0: s(j) = 0 : NEXT j

    FOR i = 1 TO n
    IF mark(ii, i) > 3 THEN mark(ii, i) = 0
    n(mark(ii, i)) = n(mark(ii, i)) + 1
    x(mark(ii, i)) = x(mark(ii, i)) + trait(i, jj)
    xxx(mark(ii, i), n(mark(ii, i))) = trait(i, jj)
    s(mark(ii, i)) = s(mark(ii, i)) + trait(i, jj) ^ 2
    NEXT i
    x = 0: s = 0
    FOR i = 1 TO ww : x1(i) = 0: s1(i) = 0 : NEXT i
    FOR i = 1 TO ww : IF n(i) < 2 THEN 300
    NEXT i

    nn = n(1) + n(2) + n(3)
    FOR i = 1 TO ww
    FOR j = 1 TO n(i)
    x = x + xxx(i, j) : s = s + xxx(i, j) ^ 2
    x1(i) = x1(i) + xxx(i, j) : s1(i) = s1(i) + xxx(i, j) ^ 2
    NEXT j, i
    sss = s - x ^ 2 / nn :sst = 0
    FOR i = 1 TO ww : sst = sst + x1(i) ^ 2 / n(i) : NEXT i
    sst = sst - x ^ 2 / nn : sse = sss - sst
    dfe = nn - ww:dft = ww -1 : f(jj) = (sst / dft) / (sse / dfe)

    NEXT jj

    qq = chr(ii)
    IF qq <> ppTHEN PRINT #4, USING "## "; chr(ii); ELSE PRINT #4, "  ";
    IF qq <> ppTHEN PRINT USING "## "; chr(ii); ELSE PRINT "  ";
    PRINT #4, pro$(ii); : PRINT #4, TAB(14); USING "## ###"; ww - 1; dfe;
    PRINT #4, TAB(22); : PRINT pro$(ii);
    PRINT TAB(14); USING "## ###"; ww - 1; dfe; : PRINT TAB(22);
    FOR i = 1 TO c
    PRINT #4, USING "####.## "; f(i); : PRINT USING "####.## "; f(i);
    NEXT i : PRINT #4, : PRINT
    pp = qq
```

```
300 NEXT ii
   PRINT #4, bb $ :PRINT bb $
   CLOE

   END
```

此程序运行后给出方差分析的自由度及F值，有关结果同时存入顺序文件 "QTLVAR" 中供DOS 的EDIT 命令及其它编辑软件调用。

运行结果：

One-way analysis of variance among alternative marker genotypes

Chr	Marker			1	2	3	4	5
5	RG573	2	73	0.56	2.48	4.15	4.83	0.66
	RG13	2	72	1.37	0.81	3.26	1.09	0.05
	RZ70	2	73	0.48	3.75	5.24	6.89	1.48
	RG697	2	64	0.67	1.66	0.14	0.23	5.45
	RG474	2	72	1.93	1.67	0.78	2.18	0.20
	RG435	2	72	0.83	1.00	0.20	0.29	2.61
	RZ225	2	72	0.06	3.52	0.94	0.20	1.84

A1.4 矩估计法 (QTLMS.BAS)

程序名称 ： QTLMS.BAS

程序用途 ： 利用 Luo和Wooliams (1993)和徐云碧(1994a)发展的以单一标记为基础的矩估计法，确定标记与 QTL 之间的重组率，并利用 LOD 值进行连锁检验。在检验连锁的同时估计 QTL 基因型的数量性状均值和方差。利用这些均值和方差，读者可进一步估计 QTL 的基因加性效应、显性效应及显性度等。

程序：

```
10  ' Program for detection of linkage between markers and
    ' and quantitative trait loci by the moment method
    ' described by Luo and Woolliams (1993) and Xu (1994a)
```

```
' This program uses data files set by the program DATASET.BAS

PRINT "Model type for variances among QTL genotypes"
INPUT "model $ = homo or hetero ?", model $

qq $ = "##.## #####.### ####.### ####.### ####.## ####.##"
PP $ = "  ####.## #####.### #####.### #####.###"
dd = 50

OPEN "traitdat" FOR INPUT AS #3
INPUT #3, pop $, p1 $, p2 $, homp1, homp2, hetero, missing
INPUT #3, m, n, C
IF pop $ = "F2" OR pop $ = "f2" THEN ww = 3 ELSE ww = 2

40 DIM trait(C, n), trait $ (C), pro $ (m), xxx(3, n)
   DIM mark(m, n), chr(m), n(3), x(3), s(3), r1(dd + 1)
   DIM mm(3), d(3, 3), e(3), ss(3), qq(3), m5(ww, C), q5(ww, C)
   DIM l(dd + 1), r3(C), rr3(C), cc3(C), qq2(3), l0(C)
   DIM df(200), dd1(20), dd2(20), mm1(3), mm2(3), lod2(C)
   DIM r4(n), mr(C), ser(C), rj(n), tt(C), r5(C)
   DIM alfa(ww, n, ww), fmi(ww, n), mm5(ww), sm5(ww), ms5(ww), ss5(ww)
   DIM m6(ww, 50), q6(ww, 50), r(3, 3)
   DIM w(3, 3, n), t(3, n), mmm(3, 50), qqq(3, 50)

70 FOR i = 1 TO c: INPUT #3, trait $ (i): NEXT i
   FOR j = 1 TO n : FOR i = 1 TO c : INPUT #3, trait(i, j)
   NEXT i, j
   CLOSE #3

   OPEN "markdat" FOR INPUT AS #3
   FOR i = 1 TO m : INPUT #3, chr(i), pro $ (i)
   FOR j = 1 TO n : INPUT #3, mark(i, j)
   NEXT j, i
   CLOSE #3

   OPEN "QTLMS" FOR OUTPUT AS #4

  bb $ = "------------------------------------------------------------------------"
  PRINT :PRINT :PRINT #4, :PRINT #4,
  PRINT "Detection of linkage between marker and quantitative traits by the"
  PRINT "moment method described by Luo and Woolliams (1993) and Xu (1994a)"
```

```
IF model$ = "homo" THEN PRINT " Homo"; ELSE PRINT " Hetero";
PRINT "scedastic model" : PRINT
PRINT #4, "Moment estimation of recombinant fraction between markers and QTLs "
PRINT #4, "described by Luo and Wooliams (1993) and Xu (1994a)"
IF model$ = "homo" THEN PRINT #4, " Homo"; ELSE PRINT #4, " Hetero";
PRINT #4, "scedastic  model" : PRINT #4,

FOR ii = 1 TO m

PRINT " Marker "; ii  : PRINT #4, " Marker "; ii
PRINT bb$
PRINT "Trait r      L      L0      LOD      m1      m2      m3      var1      var2      var3"
PRINT bb$ : PRINT #4, bb$
PRINT #4, "Trait r      L      L0      LOD      m1      m2      m3      var1      var2      var3"
PRINT #4, bb$

FOR jj = 1 TO c

mm =0: q1 = 0: n0 = 0
FOR i = 1 TO n
IF mark(ii, i) <> 0 THEN mm = mm + trait(jj, i): n0 = n0 + 1
NEXT i : mm =mm / n0
FOR i = 1 TO n
IF mark(ii, i) <> 0 THEN q1 = q1 + (trait(jj,    i) - mm) ^ 2
NEXT i : q1 = q1 / n0
l0 = 0
FOR i = 1 TO n
IF mark(ii, i) = 0 THEN 50
L00 = EXP(-(trait(jj,   i) - mm) ^ 2 / (2 * q1)) / SQR(2* 3.1415926# * q1)
l0 = l0 + LOG(L00) / LOG(10)
50  NEXT i
l0(jj)  = l0 : n4 = 0
FOR j = 0 TO ww : n(j) = 0: x(j) = 0: s(j) = 0 : NEXT j
FOR i = 1 TO n
n(mark(ii, i)) = n(mark(ii, i)) + 1
x(mark(ii, i)) = x(mark(ii, i)) + trait(jj,    i)
xxx(mark(ii, i), n(mark(ii, i))) = trait(jj,    i)
s(mark(ii, i)) = s(mark(ii, i)) + trait(jj,    i) ^ 2
NEXT i
FOR i = 1 TO ww
IF n(i) < 3 THEN PRINT : GOTO 3000
```

```
    NEXT i
    FOR j = 1 TO ww
    s(j) = (s(j) - x(j) ^ 2 / n(j)) / (n(j) - 1) : x(j) = x(j) / n(j)
    NEXT j
    FOR i = 1 TO ww : mm(i) = x(i)
    IF model$ = "homo" THEN qq(i) = q1 ELSE qq(i) = s(i)
    NEXT i

190 FOR kk = 0 TO dd

    ttt = 0
    IF kk = 0 THEN r = 0 ELSE r = r + .01
    r1(kk) = r

1500 IF ww = 3 THEN 200
    r(1, 1) = 1 - r: r(1, 2) = r: r(2, 1) = r: r(2, 2) = 1 - r
    mm(1) = ((1 - r) * x(1) - r * x(2)) / (1 - 2 * r)
    mm(2) = (-r * x(1) + (1 - r) * x(2)) / (1 - 2 * r)
    GOTO 300

200 r(1, 1) = (1 - r) ^ 2: r(1, 2) = 2 * r * (1 - r): r(1, 3) = r ^ 2
    r(2, 1) = r * (1 - r): r(2, 2) = 1 - 2 * r + 2 * r ^ 2
    r(2, 3) = r * (1 - r)
    r(3, 1) = r ^ 2: r(3, 2) = 2 * r * (1 - r): r(3, 3) = (1 - r) ^ 2
    aa = (1 - 2 * r) ^ 2 : bb = 2 * r * (1 - r)
    cc = (1 - r) ^ 2 + r ^ 2
    mm(1) = ((1 - r) ^ 2 * x(1) - bb * x(2) + r ^ 2 * x(3)) / aa
    mm(2) = (r * (r - 1) * x(1) + cc * x(2) + r * (r - 1) * x(3)) / aa
    mm(3) = (r ^ 2 * x(1) - bb * x(2) + (1 - r) ^ 2 * x(3)) / aa

300 ttt = ttt + 1
    FOR i = 1 TO ww
    FOR j = 1 TO ww : d(i, j) = (mm(j) - x(i)) ^ 2
    NEXT j, i
    FOR i = 1 TO 3: e = 0
    FOR j = 1 TO 3 : e = e + r(i, j) * d(i, j) : NEXT j
    e(i) = s(i) - e
    NEXT i
    IF ww = 3 THEN 400
    qq(1) = ((1 - r) * e(1) - r * e(2)) / (1 - 2 * r)
    qq(2) = (-r * e(1) + (1 - r) * e(2)) / (1 - 2 * r)
    qq = (qq(1) * n(1) + qq(2) * n(2)) / (n(1) + n(2))
```

```
    IF model$ = "homo" THEN qq(1)= qq: qq(2) = qq
    GOTO 500

400 qq(1) = ((1 - r) ^ 2 * e(1) - bb * e(2) + r ^ 2 * e(3)) / aa
    qq(2) = (r * (r - 1) * e(1) + cc * e(2) + r * (r - 1) * e(3)) / aa
    qq(3) = (r ^ 2 * e(1) - bb * e(2) + (1 - r) ^ 2 * e(3)) / aa
    qq = (qq(1) * n(1) + qq(2) * n(2) + qq(3) * n(3)) / (n(1) + n(2) + n(3))
    IF model$ = "homo" THEN qq(1)= qq: qq(2) = qq: qq(3) = qq

500 FOR i = 1 TO ww
    IF qq(i) <= .0000001# OR qq(i) = 0 THEN yy= kk - 1: GOTO 1950
    NEXT i
    r1 = 0
    FOR j= 1 TO ww:r2 = 0
    FOR i = 1 TO n(j): rr = 0
    FOR k =1 TO ww
    aq1 = EXP(-(xxx(j, i) - mm(k)) ^ 2 / (2 * qq(k))
    aq2 = SQR(2* 3.1415926# * qq(k))
    rr = rr + r(j, k) * aq1 / aq2
    NEXT k
    IF rr < 1D-42 THEN rr = .000001#
    r2 = r2 + LOG(rr) / LOG(10)
    NEXT i : r1 = r1 + r2
    NEXT j
    FOR i = 1 TO ww : m6(i, kk) = mm(i): q6(i, kk) = qq(i) : NEXT i
    l(kk) = r1: r1(kk) = r : yy = kk

1940 NEXT kk

1950 rr = l(yy): r = r1(yy)
     FOR kk = 0 TO yy
     IF l(kk) > rr OR l(kk) = rr THEN 1955 ELSE 1958
1955 rr = l(kk): r = r1(kk)
     FOR i = 1 TO ww
     m5(i, jj) = m6(i, kk): q5(i, jj) = q6(i, kk)
     NEXT i
1958 NEXT kk

1960 r5(jj) = r: lod2(jj) = rr

     PRINT USING "##"; jj;
     PRINT TAB(4);
```

```
        PRINT USING qq$; r5(jj); lod2(jj);  l0(jj);  lod2(jj) - l0(jj);  m5(1, jj);  m5(2, jj);
        IF ww=3 THEN PRINT USING PP$; m5(3, jj); q5(1, jj);  q5(2, jj);  q5(3, jj)
        IF ww=2 THEN PRINT USING PP$; 0; q5(1, jj); q5(2, jj)
        PRINT #4, USING "##"; jj; : PRINT #4, TAB(4);
        PRINT #4, USING qq$; r5(jj); lod2(jj); l0(jj);  lod2(jj)- l0(jj);  m5(1, jj);  m5(2, jj);
        IF ww=3 THEN PRINT #4, USING PP$; m5(3, jj); q5(1, jj); q5(2, jj);  q5(3, jj)
        IF ww=2 THEN PRINT #4, USING PP$; 0; q5(1, jj); q5(2, jj)

2000 NEXT jj
        PRINT bb$: PRINT #4, bb$

3000 NEXT ii
        CLOSE

        END
```

本程序可根据需要处理两种不同的方差模型：QTL 基因型间数量性状方差同质模型和方差异质模型。在程序运行时，首先要求选择其中一种模型。对于方差同质模型，输入 "homo"，对于方差异质模型，输入 "hetero"。两种模型下均给出标记与 QTL 之间的重组率、LOD 值以及 QTL 基因型的均值和方差的估计值，同时将结果存入顺序文件"QTLMS"，可供 DOS 的 EDIT 命令或其它编辑软件调用。

运行结果：

1. 方差同质模型

```
Moment estimation of recombinant fraction between markers and QTLs
described by Luo and Wooliams (1993) and Xu (1994a)
  Homoscedastic model

Marker 1
```

Trait	r	L	L0	LOD	m1	m2	m3	var1	var2	var3
1	0.00	-112.888	-113.128	0.239	113.65	111.43	111.45	56.990	56.990	56.990
2	0.00	-124.949	-126.019	1.070	115.35	115.54	109.36	118.224	118.224	118.224
3	0.00	-78.227	-79.991	1.764	23.66	24.72	22.70	6.985	6.985	6.985
4	0.11	-181.855	-184.000	2.145	190.03	259.23	185.86	2935.974	2935.974	2935.974
5	0.29	-144.321	-144.705	0.384	67.18	36.10	58.39	219.940	219.940	219.940

2. 方差异质模型

Moment estimation of recombinant fraction between markers and QTLs described by Luo and Wooliams (1993) and Xu (1994a)
Heteroscedastic model

Marker 1

Trait	r	L	L0	LOD	m1	m2	m3	var1	var2	var3
1	0.00	-112.833	-113.128	0.294	113.65	111.43	111.45	58.868	51.919	64.069
2	0.00	-124.724	-126.019	1.295	115.35	115.54	109.36	136.118	126.811	89.957
3	0.00	-77.077	-79.991	2.914	23.56	24.72	22.70	11.338	6.815	3.908
4	0.06	-181.936	-184.000	2.064	197.43	251.59	193.73	3805.557	3354.524	3254.589
5	0.00	-144.291	-144.705	0.414	53.64	47.09	49.95	368.709	344.534	466.023

... ...

A1.5 最大似然法 (QTLMLE.BAS)

程序名称 : QTLMLE.BAS

程序用途 : 利用徐云碧(1994a)发展的以单一标记为基础的最大似然法，确定标记与 QTL 之间的重组率，并利用 LOD 值进行连锁检验。在检验连锁的同时估计 QTL 基因型的数量性状均值和方差。利用这些均值和方差，读者可进一步估计 QTL 的基因加性效应、显性效应及显性度等。

程序:

```
10  ' Program for detection of linkage between markers and
    ' and quantitative trait loci by the maximum likelihood
    ' method described by Xu(1994a)
    ' This programuses data files set by the program DATASET.BAS
    PRINT "Model type for variances among QTL genotypes"
    INPUT "model$ = homo or hetero ?", model$

    qq$ = "##.## #####.### ####.### ####.### ####.## ####.##"
    PP$ = "  ####.## #####.### #####.### #####.###"
    dd = 50
    bb$ = "-------------------------------------------------------------------------"
```

```
    OPEN "traitdat" FOR INPUT AS #3
    INPUT #3, pop$, p1$, p2$, homp1, homp2, hetero, missing
    INPUT #3, m, n, C
    IF pop$ = "F2" OR pop$ = "f2" THEN ww = 3 ELSE ww = 2

40  DIM trait(C, n), trait$(C), pro$(m), xxx(3, n)
    DIM mark(m, n), chr(m), n(3), x(3), s(3), r1(dd + 1)
    DIM mm(3), d(3, 3), e(3), ss(3), qq(2), m5(ww, C), q5(ww, C)
    DIM l(dd + 1), r3(C), rr3(C), cc3(C), qq2(3), l0(C)
    DIM df(200), dd1(20), dd2(20), mm1(3), mm2(3), lod2(C)
    DIM r4(n), mr(C), ser(C), rj(n), tt(C), r5(C)
    DIM alfa(ww, n, ww), fmi(ww, n), mm5(ww), sm5(ww), ms5(ww), ss5(ww)
    DIM m6(ww, 50), q6(ww, 50), r(3, 3)
    DIM w(3,3, n), t(3, n), mmm(3, 50), qqq(3, 50)

70  FOR i = 1 TO c : INPUT #3, trait$(i)
    NEXT i

    FOR j = 1 TO n : FOR i = 1 TO c : INPUT #3, trait(i, j)
    NEXT i, j
    CLOSE #3

    OPEN "markdat" FOR INPUT AS #3
    FOR i = 1 TO m:INPUT #3, chr(i), pro$(i)
    FOR j = 1 TO n : INPUT #3, mark(i, j)
    NEXT j, i
    CLOSE #3

    OPEN "qtlmle" FOR OUTPUT AS #4

    PRINT :PRINT
    PRINT " Detection of linkage between marker and quantitative traits by the"
    PRINT " maximum likelihood method described by Xu (1994a)"
    IF model$ = "homo" THEN PRINT "  Homo"; ELSE PRINT "  Hetero";
    PRINT "scedastic model": PRINT :PRINT #4, :PRINT #4,
    PRINT #4, "Detection of linkage between marker and quantitative traits by the"
    PRINT #4, "Maximum likelihood estimation of recombinant fraction "
    PRINT #4, " under ";
    IF model$ = "homo" THEN PRINT #4, "homo"; ELSE PRINT #4, "hetero";
    PRINT #4, "scedastic model": PRINT #4,
```

```
    FOR ii = 1 TO m

    PRINT " Marker "; ii  : PRINT #4, " Marker "; ii
    PRINT bb$
    PRINT "Trait r      L      L0     LOD     m1      m2      m3      var1     var2     var3"
    PRINT bb$
    PRINT #4, bb$
    PRINT #4, "Trait r      L      L0     LOD     m1      m2      m3      var1     var2     var3"
    PRINT #4, bb$

    FOR jj = 1 TO c

    mm =0: q1 = 0: n0 = 0
    FOR i = 1 TO n
    IF mark(ii, i) <> 0 THEN mm = mm + trait(jj, i): n0 = n0 + 1
    NEXT i : mm =mm / n0
    FOR i = 1 TO n
    IF mark(ii, i) <> 0 THEN q1 = q1 + (trait(jj,    i) - mm) ^ 2
    NEXT i : q1 = q1 / n0

    l0 = 0
    FOR i = 1 TO n
    IF mark(ii, i) = 0 THEN 50
    L00 = EXP(-(trait(jj,   i) - mm) ^ 2 / (2 * q1)) / SQR(2* 3.1415926# * q1)
    l0 = l0 + LOG(L00)/LOG(10)
50  NEXT i
    l0(jj)  = l0 : n4 = 0
    FOR j = 0 TO ww : n(j) = 0: x(j) = 0: s(j) = 0 : NEXT j
    FOR i = 1 TO n
    n(mark(ii, i)) = n(mark(ii, i)) + 1
    x(mark(ii, i)) = x(mark(ii, i)) + trait(jj,    i)
    xxx(mark(ii, i), n(mark(ii, i))) = trait(jj,    i)
    s(mark(ii, i)) = s(mark(ii, i)) + trait(jj,    i) ^ 2
    NEXT i
    FOR i = 1 TO ww
    IF n(i) < 3 THEN PRINT : GOTO 3000
    NEXT i
    FOR j = 1 TO ww
    s(j) = (s(j) - x(j) ^ 2 / n(j)) / (n(j) - 1)
    x(j) = x(j) / n(j) : n4 = n4 + n(j)
    NEXT j
```

```
    FOR i = 1 TO ww : mm(i) = x(i)
    IF model$ = "homo" THEN qq(i) = q1 ELSE qq(i) = s(i)
    NEXT i : y$ = "n"

101 FOR kk = 0 TO dd
    ttt = 0
    IF kk = 0 THEN r = 0 ELSE r = r + .01
    r1(kk) = r

1500 IF ttt > 0 THEN 1505
    IF ww = 3 THEN 200
    r(1, 1) = 1 - r: r(1, 2) = r: r(2, 1) = r: r(2, 2) = 1 - r
    GOTO 1505

200 r(1, 1) = (1 - r) ^ 2: r(1, 2) = 2 * r * (1 - r): r(1, 3) = r ^ 2
    r(2, 1) = r * (1 - r): r(2, 2) = 1 - 2 * r + 2 * r ^ 2
    r(2, 3) = r * (1 - r)
    r(3, 1) = r ^ 2: r(3, 2) = 2 * r * (1 - r): r(3, 3) = (1 - r) ^ 2

1505 ttt = ttt + 1
    FOR i = 1 TO ww
    FOR k = 1 TO n(i): fmi = 0
    FOR j = 1 TO ww
    aq1 = EXP(-(xxx(i, k) - mm(j)) ^ 2 / (2 * qq(j))
    aq2 = SQR(2* 3.1415926# * qq(j))
    fmi = fmi + r(i, j) * aq1/aq2
    NEXT j
    fmi(i, k) = fmi
    NEXT k, i

    FOR i = 1 TO ww
    FOR k = 1 TO n(i) : FOR j = 1 TO ww
    aq1 = EXP(-(xxx(i, k) - mm(j)) ^ 2 / (2 * qq(j))
    aq2 = SQR(2* 3.1415926# * qq(j))
    alfa(i, k, j) = r(i, j) * aq1/aq2
    alfa(i, k, j) = alfa(i, k, j) / fmi(i, k)
    NEXT j, k
    NEXT i

    FOR j = 1 TO ww: mmj1 = 0
    FOR i = 1 TO ww : FOR k = 1 TO n(i)
    mmj1 = mmj1 + alfa(i, k, j) * xxx(i, k)
```

```
    NEXT k, i
    mmj2= 0
    FOR i = 1 TO ww
    FOR k =1 TO n(i) : mmj2= mmj2+ alfa(i, k, j)
    NEXT k, i: mm(j) = mmj1/ mmj2
    NEXT j

    IF model$ = "homo" THEN 1590

    FOR j= 1 TO ww:qq1 = 0
    FOR i = 1 TO ww
    FOR k =1 TO n(i)
    qq1 = qq1 + alfa(i, k, j) * (mm(j) - xxx(i, k)) ^ 2
    NEXT k, i: mmj2= 0
    FOR i = 1 TO ww
    FOR k =1 TO n(i)
    mmj2= mmj2+ alfa(i, k, j)
    NEXT k, i: qq(j) = qq1 / mmj2
    NEXT j
    GOTO 1600

1590 qq1 = 0
    FOR j= 1 TO ww
    FOR i = 1 TO ww : FOR k =1 TO n(i)
    qq1 = qq1 + alfa(i, k, j) * (mm(j) - xxx(i, k)) ^ 2
    NEXT k, i, j
    FOR j= 1 TO ww : qq(j) = qq1 / n4 : NEXT j

1600 FOR i = 1 TO ww
    IF qq(i) < .00001# OR qq(i) = 0 THEN rr =1(kk - 1): r = r1(kk - 1): GOTO 1950
    NEXT i

    r1 = 0
    FOR j= 1 TO ww:r2 = 0
    FOR i = 1 TO n(j): rr = 0
    FOR k =1 TO ww
    aq1 = EXP(-(xxx(j, i) - mm(k)) ^ 2 / (2 * qq(k))
    aq2 = SQR(2* 3.1415926# * qq(k))
    rr = rr + r(j, k) * aq1 / aq2
    NEXT k
    r2 = r2 + LOG(rr)/ LOG(10)
    NEXT i : r1 = r1 + r2
```

```
    NEXT j

    IF ttt = 1 THEN r0 = r1: GOTO 1500
    IF r1 - r0 > .001 AND ttt < 200 THEN r0 = r1: GOTO 1500
    l(kk) = r1: r1(kk) = r
    FOR i = 1 TO ww : m6(i, kk) = mm(i): q6(i, kk) = qq(i) : NEXT i
    yy = kk

    NEXT kk

1950 rr = l(yy): r = r1(yy)
    FOR kk = 0 TO yy
    IF l(kk) > rr OR l(kk) = rr THEN 1955 ELSE 1958
1955 rr = l(kk): r = r1(kk)
    FOR i = 1 TO ww : m5(i, jj) = m6(i, kk): q5(i, jj) = q6(i, kk) : NEXT i
1958 NEXT kk

1960 r5(jj) = r: lod2(jj) = rr
    PRINT USING "###"; jj; : PRINT TAB(5);
    PRINT USING qq $ ; r5(jj); lod2(jj); l0(jj); lod2(jj) - l0(jj); m5(1, jj); m5(2, jj);
    IF ww =3 THEN PRINT USING PP $ ; m5(3, jj); q5(1, jj); q5(2, jj); q5(3, jj)
    IF ww =2 THEN PRINT USING PP $ ; 0; q5(1, jj); q5(2, jj)
    PRINT #4, USING "###"; jj; : PRINT #4, TAB(5);
    PRINT #4, USING qq $ ; r5(jj); lod2(jj); l0(jj); lod2(jj) - l0(jj); m5(1, jj); m5(2, jj);
    IF ww =3 THEN PRINT #4, USING PP $ ; m5(3, jj); q5(1, jj); q5(2, jj); q5(3, jj)
    IF ww =2 THEN PRINT #4, USING PP $ ; 0; q5(1, jj); q5(2, jj)

2000 NEXT jj
    PRINT bb $ : PRINT #4, bb $

3000 NEXT ii
    CLOSE

    END
```

本程序可根据需要处理两种不同的方差模型，QTL 基因型间数量性状方差同质模型和方差异质模型。 在程序运行时，首先要求选择其中一种模型。对于方差同质模型， 输入 "homo"，对于方差异质模型，输入 "hetero"。两种模型下均给出标记与 QTL之间的重组率、LOD 值以及QTL基因型的均值和方差的估计值，同时将结果存入顺序文件"QTLMLE"，供 DOS 的 EDIT命令或其它编辑软件调用。

运行结果：

1. 方差同质模型

Detection of linkage between marker and quantitative traits by the Maximum likelihood estimation of recombinant fraction under homoscedatic model

Marker 1

Trait	r	L	L0	LOD	m1	m2	m3	var1	var2	var3
1	0.44	-112.702	-113.128	0.426	114.82	114.09	104.24	36.996	36.996	36.996
2	0.00	-124.936	-126.019	1.083	115.35	115.54	109.36	113.581	113.581	113.581
3	0.16	-78.000	-79.991	1.991	21.79	25.81	22.35	4.009	4.009	4.009
4	0.10	-181.835	-184.000	2.165	190.42	258.08	187.96	2879.268	2879.268	2879.268
5	0.34	-143.638	-144.705	1.072	57.65	35.46	71.70	143.479	143.479	143.479

... ...

2. 方差异质模型

Detection of linkage between marker and quantitative traits by the Maximum likelihood estimation of recombinant fraction under heteroscedatic model

Marker 1

Trait	r	L	L0	LOD	m1	m2	m3	var1	var2	var3
1	0.43	-111.957	-113.128	1.170	119.98	111.61	105.06	15.829	16.151	62.612
2	0.00	-124.709	-126.019	1.310	115.35	115.54	109.36	128.111	123.383	85.868
3	0.00	-77.062	-79.991	2.928	23.66	24.72	22.70	10.671	6.631	3.730
4	0.14	-179.537	-184.000	4.463	186.39	264.78	179.71	1247.771	3715.552	991.904
5	0.39	-141.939	-144.705	2.766	55.07	35.71	77.70	53.305	159.750	24.051

... ...

A1.6 区间作图法 (QTLINT.BAS)

程序名称 : QTLINT.BAS, 此程序等价于徐云碧等 (1995a) 的 QTLMAP-I。

程序用途 : 利用Lander和Botstein(1989)、van Ooijen(1992)和徐云碧等(1994a)发展的以两个相邻标记为基础的区间作图法，确定标记与QTL之间的重组率及标记区间上LOD值的大小。

DATA 语句:

除了调用 "Traitdat" 和 "Markdat" 两个顺序文件外，区间作图还要求提供用于分析的标记间的连锁关系和遗传距离 (cM)。在DATA语句中，600句为用于分析的染色体数。从 610语句开始为各个染色体的名称及该染色体所属连锁群所含的遗传标记数，随后是该连锁群上的第1标记名称、第1标记与第2标记之间的遗传图距 (cM)、第2标记名称、第2标记与第3标记之间的遗传图距，...； 第2染色体的名称及其标记数，...； 第chr 染色体的名称及其标记数，...。此外，还应根据所分析的物种的染色体数目，改变第15语句中ch的赋值，根据单个连锁群所含的最大标记数改变18语句中m1的赋值。

程序:

```
10  ' The program for QTL mapping by two franking markers using procedures
    ' described by Lander and Botstein (1989), van Ooijen (1992) and Xu et
    ' al. (1995a)
    ' This program uses data files set by the program DATASET.BAS
15  ' ch = number of chromosome
    ch = 12
    ' m1 = the maximum of marker number in one chromosome
18  m1 = 50
    p1$ = "### ###    ####.#### ####.#### ####.######"
    p2$ = "### ###    ####.#### ####.#### ####.######"
20  OPEN "Traitdat" FOR INPUT AS #3

    INPUT #3, pop$, p1$, p2$, hompl, homp2, hetero, missing
    INPUT #3, m, nn, c
    IF pop$ = "F2" OR pop$ = "f2" THEN tt = 3 ELSE tt = 2

    DIM trait(nn, c), trait$(c), pro$(m), m(ch), m$(ch, m1), f(ch, m1)
    DIM mark(m, nn), chr(m), ch(ch), a(3), b(3), al(3), ch$(ch)
    DIM w(3, 3, 3), mm(9), ww(nn, 3), w1(nn, 3), Y(nn), m1(nn), m2(nn)

30  FOR i = 1 TO c : INPUT #3, trait$(i) : NEXT i
    FOR i = 1 TO nn
    FOR j = 1 TO c : INPUT #3, trait(i, j)
```

```
    NEXT j, i
    CLOSE

40  OPEN "markdat" FOR INPUT AS #3
    FOR i = 1 TO m: INPUT #3, chr(i), pro$(i)
    FOR j = 1 TO nn: INPUT #3, mark(i, j)
    NEXT j, i
    CLOSE
    mmm = 0

    OPEN "QTLINT" FOR OUTPUT AS #4
50  READ a$, chr
    PRINT a$; " (n="; chr; ")": PRINT
    PRINT #4, a$; " (n="; chr; ")": PRINT #4,
    FOR i = 1 TO chr: READ ch$(i), m(i)
    PRINT ch$(i), "    (Number of marker, m="; m(i); ")"
    PRINT :PRINT TAB(4);
    PRINT #4, ch$(i), "    (Number of marker, m="; m(i); ")"
    PRINT #4, :PRINT #4, TAB(4);
    FOR j = 1 TO m(i) - 1
    READ m$(i, j), f(i,   j)
    PRINT m$(i, j); "-"; f(i,   j); "-";
    PRINT #4, m$(i, j); "-"; f(i,   j); "-";
    mmm = mmm + f(i, j)
    NEXT j
    READ m$(i, m(i))
    PRINT #4, m$(i, m(i)): PRINT #4, :PRINT m$(i, m(i)): PRINT
    NEXT i
    PRINT "Total length of genome ="; mmm; "cM": PRINT
    PRINT #4, "Total length of genome ="; mmm; "cM": PRINT #4,

60  OPEN "QTLINT-B" FOR OUTPUT AS #1
    WRITE #1, m, nn, c: WRITE #1, a$, chr

    FOR cc= 1 TO c

    WRITE #1, cc
    PRINT "Character "; cc; : PRINT ":"; trait$(cc)
    PRINT #4, "Character "; cc; : PRINT #4, ":"; trait$(cc)

    FOR ii = 1 TO chr
```

```
    WRITE #1, ch $ (ii): PRINT ch $ (ii): PRINT #4, ch $ (ii)

    FOR kk = 1 TO m(ii) - 1

    PRINT TAB(4); "Markers: ";
    PRINT TAB(4); m $ (ii, kk);: PRINT "-";
    PRINT USING "###.#"; f(ii, kk); : PRINT "cM"; "-"; m $ (ii, kk + 1)
    PRINT #4, TAB(4); "Markers: ";
    PRINT #4, TAB(4); m $ (ii,kk); : PRINT #4, "-";
    PRINT #4, USING "###.#"; f(ii, kk); : PRINT #4, "cM"; "-"; m $ (ii, kk + 1)
    WRITE #1, kk, m $ (ii, kk), f(ii,  kk)

    FOR j = 1 TO m
    IF pro $ (j) = m $ (ii, kk) THEN tt1 = j
    NEXT j
    FOR j = 1 TO m
    IF pro $ (j) = m $ (ii, kk + 1) THEN tt2 = j
    NEXT j

    n = 0
    FOR i = 1 TO nn
    IF mark(tt1, i) = 0 OR mark(tt2, i) = 0 THEN 70
    IF mark(tt1, i) > tt OR mark(tt2, i) > tt THEN 70
    n = n + 1: Y(n) = trait(i,  cc)
    m1(n) = mark(tt1, i): m2(n) = mark(tt2, i)
70  NEXT i

    a0 = 0
    FOR j = 1 TO n: a0 = a0 + Y(j): NEXT j
    a0 = a0 / n
    q0 = 0
    FOR j = 1 TO n: q0 = q0 + (Y(j) - a0) ^ 2: NEXT j
    q0 = q0 / n
    l0 = 0
    FOR j = 1 TO n
    a = EXP(-(Y(j) - a0) ^ 2 / (2 * q0)) / SQR(2* 3.1415926# * q0)
    l0 = l0 + LOG(a) / LOG(10)
    NEXT j
    WRITE #1, n, l0
    PRINT TAB(4); "N="; n
```

```
    PRINT TAB(21); " cM   cycle       LOG(L)     LOG(L0)       LOD"
    PRINT TAB(21); "-------------------------------------------------"
    PRINT #4, TAB(4); "N="; n
    PRINT #4, TAB(21); " cM   cycle       LOG(L)     LOG(L0)       LOD"
    PRINT #4, TAB(21); "-------------------------------------------------"

    map= 1
80  IF map>=f(ii,   kk) THEN 400
    a = EXP(2* map/ 100) - EXP(-2 * map/ 100)
    b = EXP(2* map/ 100) + EXP(-2 * map/ 100)
    r1 = .5 * a / b
    c1 = f(ii,   kk) / 100
    c1 = .5 * (EXP(2 * c1) - EXP(-2 * c1)) / (EXP(2 * c1) + EXP(-2 * c1))
    r2 = (c1 - r1) / (1 - 2 * r1)
    IF tt = 3 THEN 90
    d = r1 * r2 + (1 - r1) * (1 - r2)
    w(1, 1, 1) = r1 * r2 / d: w(1, 1, 2) = (1 - r1) * (1 - r2) / d
    d = r1 * (1 - r2) + r2 * (1 - r1)
    w(1, 2, 1) = r1 * (1 - r2) / d: w(1, 2, 2) = r2 * (1 - r1) / d
    d = (1 - r1) * r2 + r1 * (1 - r2)
    w(2, 1, 1) = (1 - r1) * r2 / d: w(2, 1, 2) = r1 * (1 - r2) / d
    d = (1 - r1) * (1 - r2) + r1 * r2
    w(2, 2, 1) = (1 - r1) * (1 - r2) / d: w(2, 2, 2) = r1 * r2 / d
    GOTO 100

90  a(1) = r1 ^ 2: a(2) = 2 * r1 * (1 - r1): a(3) = (1 - r1) ^ 2
    b(1) = r2 ^ 2: b(2) = 2 * r2 * (1 - r2): b(3) = (1 - r2) ^ 2

    FOR i = 1 TO 3: FOR j = 1 TO 3
    w(i, j, 1) = a(i) * b(j)
    NEXT j, i
    FOR i = 1 TO 3: FOR j = 1 TO 3
    w(3 - i + 1, 3 - j + 1, 3) = w(i, j, 1)
    NEXT j, i
    FOR i = 1 TO 3: FOR j = 1 TO 3
    w(i, j, 2) = 2 * SQR(w(i, j, 1) * w(i, j, 3))
    d = w(i, j, 1) + w(i, j, 2) + w(i, j, 3)
    w(i, j, 1) = w(i, j, 1) / d: w(i, j, 2) = w(i, j, 2) / d
    w(i, j, 3) = w(i, j, 3) / d
    NEXT j, i

100 FOR i = 1 TO n: FOR k = 1 TO tt
```

```
    ww(i, k) = w(m1(i), m2(i), k)
    NEXT k, i
    FOR k =1 TO tt: a(k) = 0: NEXT k
    FOR k =1 TO tt: p = 0
    FOR j = 1 TO n
    p = p + ww(j, k): a(k) = a(k) + ww(j, k) * Y(j)
    NEXT j: a(k) = a(k) / p
    NEXT k
    q = 0
    FOR k =1 TO tt: FOR j = 1 TO n
    q = q + ww(j, k) * (Y(j) - a(k)) ^ 2
    NEXT j, k: q = q / n
    x = 0

110 l1 = l
    FOR j = 1 TO n: w = 0
    FOR k =1 TO tt
    a = EXP(-(Y(j) - a(k)) ^ 2 / (2 * q)) / SQR(2* 3.1415926# * q)
    w = w + ww(j, k) * a
    NEXT k
    FOR k =1 TO tt
    a = EXP(-(Y(j) - a(k)) ^ 2 / (2 * q)) / SQR(2* 3.1415926# * q) / w
    w1(j, k) = ww(j, k) * a
    NEXT k, j
    FOR k =1 TO tt: a(k) = 0: NEXT k
    FOR k =1 TO tt: p = 0
    FOR j = 1 TO n
    p = p + w1(j, k): a(k) = a(k) + w1(j, k) * Y(j)
    NEXT j: a(k) = a(k) / p
    NEXT k
    q = 0
    FOR j = 1 TO n: FOR k =1 TO tt
    q = q + w1(j, k) * (Y(j) - a(k)) ^ 2
    NEXT k, j: q = q / n
    l = 0
    FOR j = 1 TO n: w = 0
    FOR k =1 TO tt
    a = EXP(-(Y(j) - a(k)) ^ 2 / (2 * q)) / SQR(2* 3.1415926# * q)
    w = w + ww(j, k) * a
    NEXT k: l = l + LOG(w) /LOG(10)
    NEXT j
    x = x + 1: IF x < 2 THEN 110
```

```
    IF x > 100 THEN 200
    IF ABS(l1 - l) > .0001 THEN 110

200 lod = l - 10
    PRINT TAB(22);
    PRINT USING p1$; map; x; l; l0; lod
    PRINT #4, TAB(22);
    PRINT #4, USING p2$; map; x; l; l0; lod
    WRITE #1, x, l
300 map= map+ 1

    GOTO 80
400 NEXT kk

    NEXT ii

    NEXT cc

500 CLOSE
    END

600 DATA "Number of Chromosome",1

610 DATA "Chromosome 5",7
    DATA RG573,16.5,RG13,14.7,RZ70,26.3
    DATA RG697,17.4,RG474,10.1,RG435,15.3,RZ225
```

本程序运行后，将给出所要分析的染色体数、每一染色体的名称及所含标记数、用于分析的标记及其相互间图距，基因组总图距；然后给出各性状每一染色体上各个标记区间内的LOD值，同时提供实际进行的迭代分析次数等信息。一般每隔 1cM 计算出 LOD 值。也可每隔 2cM 计算出 LOD 值，这时只要将 300 语句中的增量该变为2。运行结果存入顺序文件 "QTLINT"，可供 DOS 的 EDIT 命令或其它编辑软件调用。另外将 LOD 值存入顺序文件 "QTLINT-B"，供制作 QTL 似然图谱时电子表格系统 Quattro PRO 3.0 调用。

运行结果：

```
Number of Chromosome (n= 1 )

Chromosome 5     (Number of  marker, m=7 )
```

RG573- 16.5-RG13-14.7-RZ70-26.3-RG697-17.4-RG474-10.1-RG435-15.3-RZ225

Total length of genome = 100.3 cM

Character 1 : Days to heading

Chromosome 5

Markers:

RG573- 16.5cM-RG13

N= 75

cM	cycle	LOG(L)	LOG(L0)	LOD
1	7	-111.3481	-111.8503	0.502274
2	8	-111.3284	-111.8503	0.521957
3	9	-111.3103	-111.8503	0.540077
4	8	-111.2937	-111.8503	0.556618
5	9	-111.2785	-111.8503	0.571854
6	9	-111.2645	-111.8503	0.585846
7	9	-111.2516	-111.8503	0.598785
8	9	-111.2396	-111.8503	0.610764
9	9	-111.2284	-111.8503	0.621918
10	9	-111.2179	-111.8503	0.632416
11	8	-111.2081	-111.8503	0.642220
12	8	-111.1989	-111.8503	0.651466
13	8	-111.1904	-111.8503	0.659935
14	8	-111.1828	-111.8503	0.667580
15	7	-111.1763	-111.8503	0.674088
16	7	-111.1711	-111.8503	0.679260

Markers:

RG13- 14.7cM-RZ70

N= 75

... ...

A2　进一步阅读的建议

第1章

分子数量遗传学新学科的形成和发展以及进展和展望可参阅徐云碧 1993a，1995)。Peterson (1992) 有关“分子生物学时代的数量遗传学”一文与本章也是密切有关的。值得一提的是Doebley(1993) 把发育和进化与分子遗传学和数量遗传学结合起来，讨论了与之有关的一些问题。

第2章

国内外介绍经典数量遗传学的著作除了已提及的以外，可供参考的主要中文著作有马育华 (1982)、兰巨生(1982)、刘来福等(1984)、高之仁 (1986)、裴新澍 (1987)、徐静斐等(1991)以及莫惠栋有关应用数量遗传学的讲座（莫惠栋 1986,1987a，1987b）。在动物数量遗传学方面可参考吴仲贤(1977)。在国外，除了英美等代表性学派的著作外，可供参考的著作还有 Crow(1986)、 Hallauer 和 Miranda Fo (1981)、 Jain (1982)、Pollak 等 (1977)、Weir 等(1988)、Wricke和Weber (1986)。

此外，还有一些具有特色的著作或论文，如Baker(1986) 关于选择指数的专著， 由Frankel (1983) 编辑的有关杂种优势的论文集等。

第4章

关于遗传图谱的构建，有几个方面可供进一步阅读，如有关多倍体物种分子标记图谱的构建 (Sorrells 1992)、杂合亲本所衍生组合的重组率估计及连锁图谱构建 (Ritter 等 1990)、近缘物种的比较作图(Brown 1994； Ahn 和Tanksley 1993; Moore等1993; Kurata等1994)。遗传图谱构建的进展可参阅有关家鼠和人的研究报道(Dietrich等 1994; Gyapay 等1994)即可见一斑。

第5章

现已将永久性遗传群体用于遗传图谱的构建和 QTL 作图。如 DH 群体在水稻(徐吉臣 等 1994)和油菜(Ferreira等1994)连锁图谱构建中的应用， SSD和 DH群体在QTL作图中的应用(Wang 等1994; 徐云碧 1993b)。

第6～9章

有关 QTL 作图的综述和综合报道可参阅 Weller (1992)、Knapp等 (1992) 和徐云碧(1994c)以及 Helentjaris和Burr (1989)。关于似然比检验、 LOD值在连锁遗传分析中的应用可参考 Ott (1985)。

Rodolphe和lefort (1993)发展了一种 QTL 作图的多标记模型，其主要的目的在于检测影响 QTL作用的染色体片段，而不是检测单个的 QTL 或估计 QTL 的效应，因而使这种方法对遗传假设的要求较少。 Jansen和Stam (1994)发展了一种可利用亲本和 F_2 资料的

多元回归作图法。Hayashi和Ukai (1994) 将似然比检验统计量分解为两个部分，以分别用于检测 QTL的加性效应和显性效应。

利用独立的全同胞家系和在连锁平衡的假定下，Knot (1994) 描述了预测 QTL 检测效率的方法。Darvasi和Soller (1994) 研究了决定标记-QTL 连锁所需的最佳标记间隔的问题。徐云碧 (1994b)研究了群体大小对标记-QTL 重组率及遗传参数估计的影响。

在 QTL 作图的误差控制方面，Churchill和Doerge (1994) 讨论了与所有 QTL作图方法有关的一个问题，即显著 QTL 效应的临界值的确定，提出了一种估计临界值的经验方法。Mangin 等(1994) 研究了当 QTL 效应较小时似然比检验的非对称特性，构建了一种新的检验统计量，并获得了无偏的置信区间。Jansen (1994) 讨论了QTL 作图中 I 型和II型错误的控制问题，并提出了一种称为 MQM、适合于多个 QTL 的两步作图法。

第10章

Edwards 和Page (1994)通过计算机模拟评价了标记辅助的选择问题。在实际应用方面，Cho等(1994)借助于分子标记对水稻F_2和 F_6群体中的 *sd*-1基因进行了选择。

Stuber (1992)、Arus和Moreno-Gonzalez(1993)、Dudley (1993)、徐云碧(1994)综述了分子标记在植物遗传育种和数量性状改良中的应用。而Beckmann和Soller (1986) 关于“植物遗传改良中的 RFLP”一文则是一篇较为全面的综述。

第11章

人工染色体作为研究的重要工具被应用于大DNA 片段的克隆。Monaco和 Larin (1994) 综述了包括 YAC、BAC、PAC 和 MAC 在内的几种人工染色体。

人类基因组研究的结果对于其它动植基因组的研究有着直接的影响，其概况和进展可参阅中译本《人类基因组的作图与测序》(朱景德等译 1990) 以及柴建华(1992,1994)、柴建华和林云富 (1993)、王钦南等 (1993)、赵寿元 (1993)等。

在基因分离与克隆方面，主要的综述见 Giberson和Somerville (1993) 和Monaco (1994)。Parrish和Nelson (1993) 评价了一些基因鉴别方法，其中涉及依赖克隆 DNA和不依赖于克隆 DNA的多种方法。关于 QTL 克隆，Crabbe等(1994) 在一篇述评中指出，转基因技术以及其它分子方法的进展，使得有可能在人类遗传学方面采用定位克隆的方法克隆影响药物敏感性的 QTL。

国内有关研究的主要进展

在国内，与本书有关的研究主要集中于分子多态性分析和标记辅助的遗传研究方面。除了本书中已提及的外，这里仅列举以作物为材料的一部分论文供参考。有关分子多态性分析可参考郑康乐等 (1990)，朱立煌等(1992)，Q. Zhang等 (1992, 1993, 1994a)，吴志浩等 (1993)，毛龙等 (1994a)。标记辅助的主基因定位或遗传差异的检测可参阅陈受宜等 (1991)，刘蔼民等(1992)，郑康乐等 (1992)，Zhu等(1993)，Liang 等 (1994)、Q. Zhang等(1994b),毛龙等 (1994b)，梁成志等(1994)，朱立煌等(1994)，王京兆等

(1995)。QTL作图理论与应用方面的研究可参阅李维明等(1993), Hu等 (1992), X. F. Zhang等1992), Wu和 Li (1994)。

A3 练 习 题

在此我们提供一个实例供读者参考。有兴趣的读者不仿利用有关数据及附录A1提供的计算机程序进行有关QTL作图的尝试。

郑康乐等(1992)利用水稻三交组合和质量基因的研究方法,进行了水稻广亲和基因的RFLP 标记定位。三交组合中标记基因型的分离类似于回交群体,而反映广亲和性强弱的结实率为质量-数量性状。附表A3.1是重新整理的有关RFLP基因型数据及各个单株的结实率(%)。

试根据以上数据进行完成下列习题:

1. 求出三交群体结实率的均值和方差,并绘出结实率的频率分布图。

2. 计算三交群体中各RFLP 标记的等位基因频率,并绘出频率分布图。

3. 以第6染色体上的标记为例,对各 RFLP 标记的基因型分离比例进行 χ^2检验,判断是否符合1:1的理论分离比例,同时利用 χ^2进行标记连锁检验。

4. 通过比较标记基因型间数量性状的均值,判断标记与结实率之间是否存在关联,并计算有关QTL的加性效应、显性效应及其对结实变异的贡献大小。

5. 通过比较标记基因型间数量性状的方差,判断标记与结实率之间是否存在关联。

6. 采用TB法,比较高结实率群体(>85%)和低结实率(<70%)群体间标记等位基因频率的差异。

7. 利用方差分析的方法判断标记与结实率之间是否存在关联。

8. 采用矩估计法和最大似然法 (QTL基因型间方差同质和异质模型估计各标记与结实率之间的重组率,并根据 LOD 值进行显著性检验,比较不同估计方法和模型下所得结果。

9. 采用区间作图方法对第1、6、7、12 四个染色体上的各个标记区间是否存在影响结实率的QTL 进行检测,绘出第6染色体的似然图谱。

10. 采用简单回归分析和多元回归分析检测与结实率有关的 QTL。

11. 比较上述各类方法所得QTL 作图结果,并进行遗传解释,完成一篇简短的论文。

附表A3.1　Pecos/南京11//秋光三交群体各植株的RFLP基因型数据及结实率

株号	RFLP标记基因型															结实率
	染色体: 1		2	4	5	6						7		12		
	RG 462	RG 345	RG 256	RG 620	RG 13	C	RG 138	RG 64	RG 456	RG 716	RG 172	RG 173	RG 156	RG 323	RG 8	(%)
1	2	2	1	2	2	1	1	2	1	2	2	1	1	1	1	79.0
2	2	2	2	2	1	2	2	1	1	1	1	2	1	2	1	79.4
3	1	1	2	1	1	1	1	1	1	2	2	1	2	2	2	81.8
4	1	1	1	2	1	2	2	1	1	1	1	2	1	1	1	84.0
5	2	2	2	2	2	2	2	1	1	1	1	2	1	2	1	84.0
6	1	1	2	2	2	1	1	2	2	2	2	1	1	2	1	84.1
7	1	2	2	1	1	1	1	1	1	1	1	1	2	2	1	85.8
8	2	2	1	2	2	1	1	2	2	2	2	2	1	2	2	87.0
9	2	2	2	2	1	1	1	1	1	1	1	1	1	2	1	87.3
10	1	1	1	2	1	1	1	2	2	2	1	2	2	2	1	87.5
11	1	1	2	1	2	1	2	2	2	1	1	1	2	2	1	87.6
12	1	1	2	1	1	1	1	1	1	1	1	1	2	2	1	89.1
13	2	2	2	2	2	1	1	1	1	1	1	2	2	2	1	89.2
14	2	2	2	1	1	1	1	1	1	2	2	1	2	1	1	89.6
15	2	2	1	2	2	1	1	1	1	2	2	2	1	2	1	90.1
16	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	90.8
17	1	1	1	2	2	1	1	1	1	1	1	2	2	2	1	90.3
18	2	0	1	2	2	1	2	1	1	1	1	2	1	2	2	65.3
19	1	1	2	1	2	2	2	2	2	2	2	1	1	2	2	65.7
20	2	1	2	1	1	2	2	2	2	2	2	1	2	2	1	66.2
21	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	66.8
22	2	2	2	2	2	1	1	1	1	1	1	1	1	2	2	67.0
23	2	2	2	2	2	2	2	2	2	2	2	1	2	2	1	67.2
24	2	2	2	2	2	2	2	2	2	2	2	1	1	2	2	68.8
25	2	2	2	2	1	1	2	2	2	2	2	2	1	1	1	70.2
26	2	1	2	1	2	2	2	2	2	2	2	1	1	2	1	71.8
27	1	1	2	1	1	2	2	1	1	1	1	2	2	2	2	71.8
28	2	2	1	1	2	2	2	2	2	2	2	1	1	2	1	72.7
29	1	2	1	1	2	2	2	2	2	2	2	1	1	2	1	73.4
30	2	0	2	2	1	1	1	2	2	2	1	2	1	2	1	74.4
31	2	2	1	1	1	2	2	2	2	2	2	2	1	1	1	74.5
32	1	1	2	2	1	2	2	2	2	2	2	1	2	1	1	76.3
33	1	1	2	1	1	1	1	1	2	1	1	1	2	2	2	77.1
34	2	2	2	2	1	1	1	1	1	1	1	2	2	2	1	77.3
35	2	2	2	2	1	2	2	1	1	1	1	2	1	2	2	31.3
36	2	1	2	1	1	2	2	2	2	1	1	1	1	1	1	43.2
37	1	1	1	1	2	2	2	2	2	2	1	2	1	1	2	43.3
38	2	1	1	1	2	2	2	2	2	2	2	2	2	1	2	46.9
39	2	2	1	2	2	2	2	1	1	1	1	1	2	2	1	47.0
40	1	1	2	2	2	2	2	1	1	1	1	1	2	2	2	47.5
41	2	2	1	2	1	2	2	2	2	2	2	2	2	2	1	48.4
42	2	2	1	2	1	2	2	2	2	1	1	2	1	2	1	50.2
43	1	2	1	2	1	2	2	2	1	1	1	1	2	2	1	52.3
44	2	2	1	2	2	2	2	2	2	2	2	2	2	1	1	53.2
45	2	2	1	2	1	2	2	2	2	1	1	1	1	1	1	54.6
46	1	1	2	2	2	2	2	1	1	1	1	2	1	2	2	55.0
47	1	1	1	2	2	2	2	2	2	2	2	2	1	2	1	55.9
48	2	2	2	2	2	2	2	2	2	1	1	2	1	2	1	59.9
49	1	2	2	1	1	2	2	2	2	2	1	2	1	2	1	62.3
50	1	2	2	1	2	2	2	2	2	2	2	1	2	2	2	63.8
51	1	1	2	1	1	2	2	2	2	2	2	2	1	1	2	64.3
52	1	1	2	2	2	1	1	1	1	1	2	1	1	2	1	91.6
53	2	2	2	1	2	1	1	1	2	2	2	2	1	2	1	93.1
54	1	1	1	1	1	1	1	1	1	2	2	1	1	1	1	93.1
55	2	1	1	2	1	1	1	1	1	1	1	1	1	2	2	83.9

1: Pecos片段/秋光片段;　2: 南京11片段/秋光片段; 0:数据缺失

参 考 文 献

马育华，1982. 植物育种的数量遗传学基础。江苏科技出版社。

王钦南，江虎军，诸嘉佑，1993. 中国的人类基因组计划。生物工程进展 14(3):52-56。

王京兆，王斌，徐琼芳，杨代常，朱英国. 1995. 用 RAPD 方法分析水稻光敏核不育基因。遗传学报 22:53-58。

毛龙，周清，王显平，胡含，朱立煌，1994 高秆野生稻与栽培稻杂交后代的 RFLP 分析。植物学报 36:11-19。

毛龙，胡含，朱立煌，薛秀庄，1994. 应用 RAPD 分析一个抗条锈病的小麦-黑麦易位系。科学通报 39:2088-2090。

兰巨生，1982. 作物遗传参数统计法。河北人民出版社。

申宗坦，徐云碧，1992. 水稻亚种间杂种不育性的遗传分析。中国水稻科学 6:105-112.

朱衡，蒋琬如，陈美玲，王久林，雷财林，凌忠专，朱立煌，1994. 稻瘟病菌株的 DNA 指纹及其与小种致病性相互关系的研究。作物学报 20:257-263。

朱立煌，1992. 植物基因定位和基因组研究进展。见：敖世洲（主编），《基因分子生物学研究进展》，上海科学技术出版社，第179-184页。

朱立煌，徐吉臣，胡乃璧，李良材，杨振玉，1992. 一对粳稻细胞质雄性不育系与恢复系的DNA限制性片段长度的多态性分析。作物学报 18:99-106。

朱立煌，徐吉臣，陈英，凌忠专，陆朝福，徐云碧，1994. 用分子标记定位一个未知的抗稻瘟病基因。中国科学（B辑）24:1048-1052。

朱景德，周郑，张爱兰，1994.（译）人类基因组的作图与测序（美国国家科学研究委员会著）。上海科学技术出版社。

刘来福，毛盛贤，黄远樟，1984. 作物数量遗传。农业出版社。

刘蔼民，李和标，张启发，姜晓红，师素云，杨官品，1992. 水稻广亲和基因在RFLP图谱上的初步定位。华中农业大学学报 11:213-219。

李维明，吴为人，卢浩然，1993. 检测作物数量性状基因与遗传标记连锁关系的方差分析法及其应用。作物学报 19:97-102。

吴志浩，徐云碧，朱立煌，1993. 籼粳稻间限制性片段长度多态性影响因素的分析。遗传学报，20:455-461。

吴仲贤，1977. 统计遗传学，科学出版社。

陈受宜，朱立煌，洪建，陈少麟，1991. 水稻抗盐突变体的分子生物学鉴定。植物学报 33:567-573。

林德光，耿旭，1987. 多变元双列杂交育种法的统计分析。热带作物学报 8(2):1-24。

郑康乐，1991. 分子标记在作物遗传育种中的应用。江苏科学技术出版社。

郑康乐，沈波，于飞，赵成章，戚秀芳，徐星明，1990. 水稻DNA限制性片段长度多态性的初步研究。中国水稻科学 4:145-149。

郑康乐，沈波，钱惠荣，王建林，1992. 应用RFLP标记研究水稻的广亲和基因。中国水稻科学 6:145-150。

赵寿元，1993. 人类基因组。上海科技教育出版社。

莫惠栋，1986. 数量遗传研究的试验设计。江苏农学院学报 7(3):51-56。

莫惠栋，1987a. 双列资料的遗传模型分析。江苏农学院学报 8(1):59-64。

莫惠栋，1987b．三重测验杂交的遗传分析。江苏农学院学报 8(2):59-64。
莫惠栋，1993．质量-数量性状的遗传分析．Ⅰ．遗传组成和主基因基因型鉴别。作物学报 19:1-6。
莫惠栋，1993．质量-数量性状的遗传分析．Ⅱ．世代平均数和遗传方差。作物学报 19:193-200。
徐云碧，1992．分子标记在数量基因定位中的应用。遗传 14(2)：45-48。
徐云碧，1993a．发展中的遗传学分支学科——分子数量遗传学。见：遗传学基础理论问题讨论文集，北京师范大学出版社，第152-166页。
徐云碧，1993b．利用 DNA 限制性片段长度多态性标记定位水稻数量性状基因。浙江农业大学博士论文。
徐云碧，1994a．分子标记与QTL之间重组率的估计及其Monte Carlo模拟。浙江农业大学学报 20:579-586.
徐云碧，1994b．QTL作图的影响因素—群体大小。浙江农业大学学报 20:573-378。
徐云碧，1994c．分子标记在植物数量性状遗传改良中的应用。第三届全国青年作物遗传育种学术讨论会论文集。中国农业科学技术出版社，第38-49页。
徐云碧，1995．分子数量遗传学研究进展与展望。上海农学院学报(增刊)。
徐云碧，申宗坦，1991．早籼稻产量构成性状的遗传分析。浙江农业学报 3:49-54。
徐云碧，申宗坦，1992a．籼粳杂种糯性基因的异常分离及其特征。中国水稻科学 6:89-92。
徐云碧，申宗坦，1992b．水稻茎叶形态性状的简化三重测交分析。作物学报 18:344-351。
徐云碧，申宗坦，1993．早籼稻品种分蘖角度的遗传分析。浙江农业学报 5:1-5。
徐云碧，申宗坦，朱立煌，陈英，1994a．水稻形态性状与分子标记的相互关联及其检测。浙江农业学报6:1-6。
徐云碧，申宗坦，徐吉臣，朱衡，陈英，朱立煌，1994b．利用分子标记连锁图谱进行水稻数量基因的区间作图。中国科学(B辑)24:971-976。
徐云碧，申宗坦，陈英，朱立煌，1995a．QTL 区间作图的统计理论和计算机软件及其应用。作物学报21:1-8。
徐云碧，申宗坦，陈英，朱立煌，1995b．利用最大似然法进行水稻产量基因的分子作图。遗传学报 22:46-52。
徐云碧，申宗坦，陈英，朱立煌，1995c．分子标记与 QTL 之间重组率的最大似然估计及其程序设计。中国农业科学 28(印刷中)。
徐云碧，申宗坦，陈英，朱立煌，1995d，水稻籼粳杂种 F_2 群体 RFLP 标记的异常分离及其染色体分布。植物学报 37（印刷中）
徐吉臣，朱立煌，陈英，陆朝福，才宏伟，1994．用双单倍体群体构建水稻的分子连锁图。遗传学报21:205-214。
徐静斐，孙五成，程融，林毅，1991．水稻数量遗传与育种。安徽科学技术出版社。
柴建华，1992．人类基因组研究路线、方法及进展。见：赵世洲（主编），基因分子生物学研究进展，上海科学技术出版社，第168-178页。
柴建华，1994．人类基因组研究新进展。生命科学 6(5):31-33。
柴建华，林云富，1993．人类基因组图谱分析进展。生物工程进展 14(3):14-24。
高之仁，1986．数量遗传学。四川大学出版社。
梁成志，童哲，宋文源，朱立煌，1994．光敏核不育基因的RFLP 分析。植物学报 36:664-670。
裴新澍，1987．数理遗传与育种，上海科学技术出版社。
Adams, W. T. 1983. Application of isozymes in tree breeding. In: S. D. Tanksley and T. S. Orton (eds) Isozymes in Plant Genetics and Breeding, Elsevier Science Publishers, Amsterdam, pp381-400.

Ahn,S. and S. D. Tanksley. 1993. Comparative linkage maps of the rice and maize genomes. Proc. Natl. Acad. Sci. USA 90:7980-7984.

Allard, R. W. 1956. Formulas and tables to facilitate the calculation of recombinantion values in heredity. Hilgardia 24:235-278.

Allard, R. W. 1960. Principles of Plant Breeding. John Wiley & Sons, Inc., New York.

Arus,P. 1988. Genetic purity of commercial seed lots. In:S. D. Tanksley and T. J. Orton (eds) Isozymes in Plant Genetics and Breeding, Elsevier Science Publishers, Amsterdam,pp415-424.

Asins, M. J. and E. A. Carbonell. 1988. Detection of linkage between restriction fragment length polymorphism markers and quantitative traits. Theor. Appl. Genet.76:623-626.

Bailey, D. C. 1983. Isozyme variation and plant breeder's rights. In: S. D. Tanksley and T. S. Orton (eds) Isozymes in Plant Genetics and Breeding, Elsevier Science Publishers, Amsterdam, pp425-440.

Baker, R. J. 1986. Selection Indices in Plant Breeding. CRC Press, Canada.

Beavis, W. D., D. Grant, M. Albertsen and R. Fincher. 1991. Quantitative trait loci for plant height in four maize populations and their associations with qualitative genetic loci. Theor. Appl. Genet. 83:141-145.

Becker, W. A. (Ed.) 1984. Manual of Quantitative Genetics. Academic Enterprises, Pullman, Washington.

Beckmann, J. S. and M. Soller. 1983. Restriction fragment length polymorphisms in genetic improvement: moethodologies,mapping and costs. Theor. Appl. Genet. 67:35-43.

Beckmann, J.S. and M. Soller. 1986. Restriction fragment length polymorphisms and genetic improvement of agricultural species. Euphytica 35:111-124.

Beckmann, J.S. and M. Soller.1986. Restriction fragment length polymorphisms in plant genetic improvement. Oxford Surveys of Plant Molecular & Cell Biology 3:196-250.

Beckmann, J. S. and M. Soller. 1988. Detection of linkage between marker loci and loci affecting quantitative traits in crosses between segregating populations. Theor. Appl. Genet. 76:228-236.

Bent, A. F., B. N. Kunkel, D. Dahlbeck, K. L. Brown, R. Schmidt, J. Giraudat, J. Leung and B. J. Staskawicz. 1994. *RPS2* of *Arabidopsis thaliana* represents a new class of plant disease resistance genes. Science 265:1856-1860.

Bonierbale, M. W., R. L. Plaisted and S. D. Tanksley. 1988. RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120: 1095-1103.

Bonierbale, M. W., R. L. Plaisted, O. Pineda, S. D. Tanksley. 1994. QTL analysis of trichrome-mediated insect resistance in potato. Theor. Appl. Genet. 87:973-987.

Botstein, D. R., R. L. White, M. Skolnick and R. W. Davis.1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am.J. Hum. Genet. 32:314-331.

Brown,S. D. M.1994. Integrating maps of the mouse genome. Current Opinion in Genetics and Development 4:389-394.

Brown, A. H. D. and B. S. Weir. 1983. Measuring genetic variability in plant populations. In:

S. D. Tanksley and T. J. Orton (eds) Isozymes in Plant Genetics and Breeding, Elsevier Science Publishers, Amsterdam, pp219-240.

Bubeck, D. M., M. M. Goodman, W. D. Beavis and D. Grant. 1993. Quantitative trait loci controlling resistance to gray leaf spot. Crop Sci. 33:838-847.

Bucio Alanis, L. 1966. Environmental and genotype-environmental components of variability: I. Inbred lines. Heredity 21:387-397.

Burke, D. T., G. F. Carle and M. V. Olson. 1987. Cloning of large segments of exogeneous DNA into yeast by means of artificial chromosome vectors. Science 236:806-812.

Burr, B. and F. A. Burr. 1991. Recombinant inbreds for molecular mapping in maize: theoretical and practical considerations. TIG 7:55-60

Burr, B., F. A. Burr, K. H. Thompson, M. C. Albertson and C. W. Stuber. 1988. Genemapping with recombinant inbreds in maize. Genetics 118:519-526.

Burr, B. et al. 1992. Pinning down loose ends: mapping telomeres and factors affecting their length. The Plant Cell 4:953-960.

Carle, G. F., M. Frank, M. V. Olson. 1986. Electrophoretic separations of large DNA molecules by periodic inversion of the electric field. Science 232:65-68.

Carbonell, E. A., T. M. Gerig, E. Balansard and M. J. Asins. 1992. Interval mapping in the analysis of nonadditive quantitative trait loci. Biometrics 48:305-315.

Carbonell, E. A., M. J. Asins, M. Baselga, E. Balansard and T. M. Gerig. 1993. Power studies in the estimation of genetic parameters and the localization of quantitative trait loci for backcross and doubled haploid population. Theor. Appl. Genet. 86:411-416.

Castle, W. E. 1921. On a method of estimating the number of genetic factors concerned in cases of blending inheritance. Science 54:93-96.

Causse, M. A., T. M. Fulton, Y. G. Cho, S. N. Ahn, J. Chunwongse, K. Wu, J. Xiao, Z. Yu, P. C. Ronald, S. E. Harrington, G. Second, S. R. McCouch and S. D. Tanksley. 1994. Saturated molecular map of the rice genome based on an inerspecific backcross population. Genetics 138:1251-1274.

Churchill, G. A. and R. W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics 138:963-971.

Churchill, G. A., J. J. Giovannoni and S. D. Tanksley. 1993. Pooling-samplingmakes high-resolution mapping practical with DNA markers. Proc. Natl. Acad. Sci. USA 90: 16-20.

Cho, Y. G., M. Y. Eun, S. R. McCouch and Y. A. Chae. 1994. The semidwarf gene, *sd*-1, of rice (*Oryza sativa* L.). II. Molecular mapping and marker-assisted selection. Theor. Appl. Genet. 89:54-59.

Choo, T. M, E. Reinbergs and K. J. Kasha. 1985. Use of haploids in breeding barkey. Plant Breed. Rev. 3:219-252.

Chu, G., D. Vollrath and R. W. Davis. 1986. Separation of large DNA molecules by contour-clamped homogeneous electric fields. Science 234:1582-1585.

Clegg, M. T. 1983. Detection and measurement of natural selection. In: S. D. Tanksley and T. J. Orton (eds) Isozymes in Plant Genetics and Breeding, Elsevier Science Publishers, Amsterdam, pp241-256.

Comstock, R. E. and R. H. Moll. 1963. Genotypic-environment interactions. Statistical Genetics

and Plant Breding. Nat. Acad. Sci. 169-196.

Cowen, N. M. 1988. The use of replicated progenies in marker-based mapping of QTL's. Theor. Appl. Genet. 75: 857-862.

Cowen, N. M. 1989. Multiple linear regression analysis of RFLP data sets used in mapping QTLs. In: T. Helentjaris and B. Burr (eds) Development and Application of Molecular Markers to Problems in Plant Genetics, Cold Spring Harbor Laboratory. pp113-116.

Crabde. J. C., J. K. Belknapand K. J. Buck. 1994. Genetic animal models of alchol and drug abuse. Science 264:1715-1723.

Crawford, D. J. 1983. Phylogenetic and systematic inferences from electrophoretic studies. In: S. D. Tanksley and T. J. Orton (eds) Isozymes in Plant Genetics and Breeding, Elsevier Science Publishers, Amsterdam, pp257-288.

Crow, J. F. 1986. Basic Concepts in Population, Quantitative Genetics, and Evolutionary Genetics. W. H. Freem& Co., New York, NY.

Darvasi, A. and M. Soller. 1992. Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. Theor. Appl. Genet. 85: 353-359.

Darvasi, A. and J. I. Weller. 1992. On the use of the moments method of estimationto obtain approximate maximum likelihood estimates of linkage between genetic marker and a quantitative locus. Heredity 68:43-46.

Darvasi, A. and M. Soller. 1994. Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. Genetics 138:1365-1373.

Darvasi, A. and M. Soller.1994. Optimum spacing of genetic markers for determining linkage between marker loci and quantitative trait loci. Theor. Appl. Genet. 89:351-357.

Darvasi, A., A. Weinreb, V. Minke.J. I. Weller and M. Soller.1993. Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics 134:943-951.

Day, N. E. 1969. Estimating the components of a mixture of normal distributions. Biometrika 56:463-474.

Dempster, A. P., N. M. Laid and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Statist. Soc. Ser. B 39:1-38.

De Vicente, M. C. and S. D. Tanksley. 1993. QTL analysis of transgressive segregation in an interspecific tomato cross. Genetics 134:585-596.

Diers, B. W., P. Keim. W. R. Fehr and R. C. Shoemaker. 1992. RFLP analysis of soybean seed protein and oil content. Theor. Appl. Genet 83: 608-612.

Dietrich, W. F., J. C. Miller, R. G. Steen, M. Merchant, D. Damron, Robert Nahf, A. Gross, D. C. Joyce, M. Wessel, R. D. Dredge, A. Marquis, L. D. Stein, N. Goodman, D. C. Page and E. S. Lander. 1994. A genetic map of the mouse with 4,006 simple sequence length polymorphisms. Nature Genetics 7:220-245.

Dodds, K. G., G. W. Montgomery and M. L. Tate. 1993. Testing for linkage between a marker locus and a major gene locus in half-sib families. J. Heredity 84:43-48.

Doebley, J. 1993. Genetics, development and plant evolution. Current Opinion in Genetics and Development 3:865-872.

Donis-Keller, H., P. Green, C. Helms, S. Cartinghour, B. Weiffendach, K. Stephens, T.P. Keith,

D. W. Bowden, D.R. Smith, E. S. Lander, D. Botstein, G. Akots, K. S. Rediker, T. Gravius, V. A. Brown, M. B. Rising, C. Parker, J. A. Powers, D. E. Watt, E. R. Kauffman, A. Bricker, P. Phippes, H. Muller-Kahle, T. R. Fulton, S. Ng, J. W. Schumm, J. C. Braman, R. G. Knowlton, D.F. Barker, S. M. Crooks, S. E. Lincoln, M. J. Daly and J. Abrahamson. 1987. A genetic linkage map of the human genome. Cell 51:319-377.

Dudley, J. W. 1992. Theory for identification of marker locus-QTL associations inpopulation by line crosses. Theor. Appl. Genet. 85.101-104.

Dudley, J. W. 1993. Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. Crop Sci. 33:660-668.

East, E. M. 1915. Studies on size inheritance in *Nictiana*. Genetics 1:164-176.

East, E. M. and H. K. Hayes. 1912. Heterozygosis in evolution and in plant breeding. Bur. Plant Ind. Bull. (U.S. A.) 243:1-58.

Edwards, M. D. and N. J. Page. 1994. Evaluation of marker-assisted selection through computer simulation. Theor. Appl. Genet. 88:376-382.

Edwards, M. D., C. W. Stuber and J. F. Wendel. 1987. Molecular marker-facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics 116:113-125.

Edwards, M. D., T. Helentjaris, S. Wright and C. W. Stuber. 1992. Molecular marker-facilitated investigations of quantitative trait loci in maize. Theor. Appl. Genet. 83: 765-774.

Ellis, T. N., D. Lee, C. M. Thomas, P. R. Simpson, W. G. Cleary, M. A. Newman and K. W. G. Burcham. 1988. 5S rRNA genes in *Pisum*: sequence, long range and chromosomal organization. Mol. Gen. Genet. 214:333-342.

Elston, R. C. 1984. The genetic analysis of quantitative trait differences between two homozygous lines. Genetics 108:733-744.

Elston, R. C. and J. Steward. 1973. The analysis of quantitative traits for simple genetic models from parental, F_1 and backcross data. Genetics 73:695-711.

Emerson, R. A. and E. M. East. 1913. The inheritance of quantitative characters in maize. Buul. Agr. Exp. Sta. Nebraska, Res. Bull. 2.

Falcolner. D. S. 1960. Introduction to Quantitative Genetics. Oliver & Boyd, Edinburg & New York.

Falcolner, D. S. 1981. Introduction to Quantitative Genetics. Longman Inc., New York.

Fatokun, C. A., D. I. Menancio-Hautea, D. Danesh and N. D. Young. 1992. Evidence for orthologous seed weight genes in cowpea and mung bean on RFLP mapping. Genetics 132:841-846.

Ferrira, M. E., P. H. Williams and T. C. Osborn. 1994. RFLP mapping of *Brassica napus* using doubled haploid lines. Theor. Appl. Genet. 89:615-621.

Fisher, R. A. 1935. The detection of linkage with dominant abnormalities. Ann. Eugen. 6:187-201.

Fisher, R. A., F. R. Immer and O. Tedin. 1932. The genetical interpretation of statistics of the third degree in the study of quantitative inheritance. Genetics 17:107-124.

Franker, R. (ed.) 1983. Heterosis, reappraisal of theory and practice. Springer-Verlag, Berlin Heidelberg.

Frei, O. M., C. W. Stuber and M. M. Goodman. 1986a. Use of allozymes as genetic markers for predicting performance in mazie single cross hybrids. Crop Sci. 26:37-42.

Frei, O. M., C. W. Stuber and M. M. Goodman. 1986b. Yield manipulation from selection on allozyme genotypes in a composite of elite corn lines. Crop Sic. 26:917-921.

Frey, K. J. and T. Horner. 1957. Heritability in standard units. Agron. J. 49:59-62.

Gamble, E. E. 1962. Gene effects in corn (*Zea mays* L.) I. separation and relative importance of gene effects for yield. Canad. J. Plant Sci. 42:339-348

Ganal, M. W. and S. D. Tanksley. 1989. Analysis of tomato DNA by pulsed field gel electrophoresis. Plant Mol. Biol. Rep. 7:17-27.

Ganal, M. W., N. D. Young and S. D. Tanksley. 1989. Pulsed field gel electrophoresis and physical mapping in the *Tm-2a* region of chromosome 9 in tomato. Mol. Gen. Genet. 215:395-400.

Ganal, M. W., M. W. Bonierbale, M. S. Roeder, W. D. Park, and S. D. Tanksley. 1991. Genetic and physical mapping of the patatin genes in potato and tomato. Mol. Gen. Genet. 225:501-509.

Ganal, M. W., P. Brounand S. D. Tanksley. 1992. Genetical mapping of tandemly repeated telomeric DNA sequences in tomato (*Lycopersicon esculentum*). Genomics 14:444-448.

Geldermann, H. 1975. Investigations on inheritance of quantitative characters in animals by gene markers. I. Methods. Theor. Appl. Genet. 46:319-330.

Gertz, A. and G. Wricke. 1989. Linkage between the incompatibility locus Z and a β-glucosidase locus in rye. Plant Breed. 102:255-259.

Gibson, S. and S. Somerville. 1993. Isolating plant genes. TIBTECH 11:306-313.

Goddard, M. E. 1992. A mixed model for analyses of data on multiple genetic markers. Theor. Appl. Genet. 83: 878-886.

Goivannoni, J. J., R. A. Wing, M. W. Ganal and S. D. Tanksley. 1991. Isolation of molecular markers from specific chromosomal interval using DNA pools from existing mapping populations. Nucl. Acid Res. 19:6553-6558.

Goldman. I. L., T. R. Rocheford and J. W. Dudley. 1993. Quantitative trait loci influencing protein and starch concertration in the Illinois Long Term Selection maize strains. Theor. Appl. Genet. 87: 217-224.

Gonella, J. A. and P. A. Peterson. 1978. Isozyme relatedness of inbred lines of maize and performance of their hybrids. Maydica 23:55-61.

Graef, G. L., W. R. Fehr and S. R. Cianzio. 1989. Relation of isozyme genotypes to quantitative characters in soybean. Crop Sci. 29:683-688.

Gray, J., S. Picton, J. Shabbear, W. Schuch and D. Grierson. 1992. Molecular biology of fruit ripening and its manipulation with antisense genes. Plant Mol. Biol. 19:69-87.

Grevelding, G., D. Becker, R. Kunze, A. V. Menges, V. Fantes, J. Schell and R. Masterson. 1992. High rates of *Ac/Ds* germinal tansposition in *Arabidopsis* suitable for gene isolation by insertional mutagenesis. Proc. Natl. Acad. Sci. USA 89:6085-6089.

Griffing, B. 1956. A generalised treatment of the use of diallel crosses in quantitative inheritance. Heredity 10:31-50

Gulick, P. J. and J. Dvorak. 1990. Selective enrichment of cDNAs from salt-stress-induced

genes in the wheatgrass, *Lophopyrum elongatum* by the formamide-phenol emulsion reassociation technique. Gene 95:173-177.

Gyapay,G., J. Morissette, A. Vignal, C. Dib, C. Fizames, P. Milasseau, S. Marc, G. Bernardi, M. Lathrop and J. Weissenbach. 1994. The 1993-94 Genethon human genetic linkage map. Nature Genetics 7:246-339.

Hadjinov, M. I., V. S. Scherbak, N. I. Banko, V. P. Gusev, T. B. Sukhorzheuskaya, and L. P. Voronova. 1982. Interrelationships between isozyme diversity and combining abilityin maize lines. Maydica 27:135-149.

Haldane, J. B. S. 1919. The combination of linkage values and the caculation of distances between the loci of linkage factors. J. Genet. 8:299-309.

Haldane, J. B. S. and C. A. B. Smith. 1947. A new estimate of the linkage between the genes for colour-blindness and haemophilia in man. Ann. Eugen. 14:10-31.

Haldane, J. B. S. and C. H. Waddington. 1931. Inbreeding and linkage. Genetics 16:357-374.

Haley, C. S. and S. A. Knott. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315-324.

Haley, C. S., S. A. Knott and J.-M. Elsen. 1994. Mapping quantitative trait lociin crosses between outbred lines using least squares. Genetics 136:1195-1207.

Hallauer, A. R. & J. B. Miranda, Fo. 1981. Quantitative Genetics in Maize Breeding. Iowa State Univ. Press.

Hanson, W. D. 1959. Early generation analysis of lengths of heterozygous chromosome segments around a locus held heterozygous with backcrossing or selfing. Genetics 44: 833-837.

Hanson, W. D. and H. F. Robinson (eds) 1963. Statistical Genetics and Plant Breeding. NAS-NRC, Washington D.C.

Harrison, B. J. and K. Mather. 1950. Polygenic variability in chromosome of *Drosophila melanogaster* obtained from the wild. Heredity 43:295-312.

Hasselbald, F. 1966. Mixture of normal distribution. Technometrics 8:231-250.

Hayashi, H., I. Czaja, H. Lubenow, J. Schell and R. Walden. 1992. Activiation of a plant gene by T-DNA tagging: auxin-independent growth in vitro. Science 258:1350-1353.

Hayashi, T and Y. Uksi. 1994. Detection of additive and dominance effects of QTLs in interval mapping of F_2 RFLP data. Theor. Appl. Genet. 87:1021-1027.

Hayes, P. M., B. H. Liu, S. J. Knapp, F. Chen, B. Jones, T. Blake, J. Franckowiak, D. Rusmusson, M. Sorrells, S. E. Ullrich, D. Wesenberg and A. Kleinhofs. 1993. Quantitative trait locus effects and environmental interaction in a sample of North American baley germplasm. Theor. Appl. Genet. 87:392-401.

Hayman, B. I. 1958. The separation of epistatic from additive and dominance variationin generation means. Heredity 12:371-390.

Heidrich-Sobrinho, E. and A. R. Cordeiro. 1975. Codominant isozymatic alleles as markers of genetic diversity correlated with heterosis in maize (*Zea mays* L.) Theor. Appl. Genet. 46:197-199.

Helentjaris, T. and B. Burr. 1989. Development and Application of Molecular Markersto to Problems in Plant Genetics. Cold Spring Harbor Laboratory.

Helentjaris, T., G. King, M. Slocum, C. Siedenstrang and S. Wegman. 1985. Restriction fragment

length polymorphisms as probes for plant diversity and their developmentas tools for applied plant breeding. Plant Mol. Biol. 5:109-118.

Helentjaris, T., D. F. Weber and S. Wright. 1986a.Use of monosomics to map cloned DNA fragments in mazie. Proc. Natl. Acad. Sci. USA 83:6035-6039.

Helentjaris, T., M. Slocum, S. Wright, A. Schaefer and J. Nienhuis.1986b. Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms.Theor. Appl. Genet. 72:761-769.

Helentjaris, T., M. A. F. Cushman and R. Winkler. 1992. Developping a genetic understanding of agronomy traits with complex inheriance. In: Y. Dettee, C. Dumas and A.Gallais (eds) Reproductive Biology and Plant Breeding. Springer-Verleg Berlin Heidelberg, pp397--406.

Heun,M. and H. R. Gregorius. 1987. A theoretical model for estimating linkage in F_2 populations with distorted single gene segregation. Biom. J. 29:397-406.

Heun,M., A. E. Kennedy, J. A. Anderson,N. L. V. Lapitan,M.E. Sorrelis and S. D. Tanksley. 1991. Construction of a restriction fragment length polymorphism map for barley (*Hordeum vulgare*). Genome 34:437-447.

Hill, A. P. 1975. Quantitative linkage: A statistical procedure for its detection and estimation. Ann. Hum. Genet. Lond. 38:439-449.

Hill, W.G. (ed.) 1984. Quantitative Genetics (I & II). Van Nostrand Reinhold Company Inc.

Hoeschele, I. and P. M. VanRaden. 1993.Bayesian analysis of linkage betweengenetic markers and quantitative trait loci. I. Prior knowledge.Theor. Appl. Genet. 85:953-960.

Hoeschele, I. and P. M. VanRaden. 1993.Bayesian analysis of linkage betweengenetic markers and quantitative trait loci. II. Combining prior knowledge with experimental evidence. Theor. Appl. Genet. 85:946-952.

Hu, Zhongli and C. W. Stuber. 1992. Methods for detection and estimation of linkage between a marker locus and quantitative trait loci. I. Using F_2 family subpopulation. J. Biomath. (Beijing) 7:12-21.

Hunter. R. B. and W. Kannenberg. 1971. Isozyme characterization of corn (*Zea mays* L.) inbreds and its relation to single cross hybrid performance. Can. J. Genet. Cytol. 13: 649-655.

Jackson, D. A. and P. R. Cook. 1985. A general method for preparing chromatin containing intact DNA. EMBO J. 4:913-918.

Jain, J. P. 1982. Statistical Techniques in Quantitative Genetics. Tata McGraw-Hill Publishing Company Limited, New Delhi.

Jansen, R. C. 1992. A general mixture model for mapping quantitative trait loci by using molecular markers. Theor. Appl. Genet. 85:252-260.

Jansen, R. C. 1993. Interval mapping of multiple quantitative trait loci. Genetics 135:205-211.

Jansen, R. C. 1994. Controlling the type I and type II errors in mapping quantitative trait loci. Genetics 138:871-881.

Jansen, R. C. and P. Stam. 1994. High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136:1447-1455.

Jayakar, S. D. 1970. On the detection and estimation of linkage between a locus influencinga

quantitative character and a marker locus. Biometrics 26:451-464.

Jayasekara, N. E. M. and J. L. Jinks. 1976. Effect of gene dispersion on estimates of components of generation means and variances. Heredity 36:31-40.

Jeffreys., A. J., V. Wilson and S. L. Thein. 1985. Individual-specific fingerprints of human DNA. Nature 316:76-79.

Jensen, J. 1989. Estimation of recombination parameters between a quantitative trait locus (QTL) and two marker gene loci. Theor. Appl. Genet. 78:613-618.

Jiang, C., X. Pan and M. Gu. 1994. The use of mixture models to detect effects of major genes on quantitative characters in a plant breeding experiment. Genetics 136:383-394.

Jung, C., M. Kleine, F. Fischer and R. G. Herrmann. 1990. Analysis of DNA from a *Beta procumbens* chromosome fragment in sugar beet carrying a gene for nematode resistance. Theor. Appl. Genet. 79:663-672.

Karp, A. and W. J. Bright. 1984. On the causes and origins of somaclonal variation. Oxford Surveys of Plant Molecular and Cell Biology 2:199-234.

Kearsey, M. J. and J L. Jinks. 1968. A general method of detecting additive, dominance and epistatic variation for metrical traits: I. Theory. Heredity 23:403-409.

Keim, P., B. W. Diers and R. C. Shoemaker. 1990a. Genetic analysis of soybean hard seededness with molecular markers. Theor. Appl. Genet. 79 :465-469.

Keim, P., B. W. Diers, T. C. Osborn and R. C. Shoemaker. 1990b. RFLP mapping in soybean-association between marker loci and variation in quantitative traits. Genetics 126:735-742.

Keitbundit, S., H. De Greve, F. Deboeck, M. VanMontagu, J.-P. Hernalsteens. 1991. *In vivo* random β-glucuronidase gene fusions in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA 88:5212-5216.

Kempthorne, O. 1957. An Introduction to Genetic Statistics. Wiley, New York.

Khush, G S. 1987. List of gene markers maintained in the Rice Genetic Stock Center, IRRI, Rice Genet. Newsl. 4:56-62.

Khush, G.S., R. J. Singh, S. C. Sur and A. L. Librojo. 1984. Primary trisomics of rice: Origin, morphology, cytology, and use in linkage mapping. Genetics 107:141-163.

Kinoshita, T. 1991. Report of the Committee on Gene Symbolization, Nomenclature and Linkage Groups. Rice Genet. Newsl. 8:2-37.

Knapp, S. J. 1991. Using molecular markers to map multiple quantitative trait loci: models for backcross, recombinant inbred, and doubled haploid progeny. Theor. Appl. Genet. 81:333-338.

Knapp, S. J. and W. C. Bridge. 1990. Using molecular markers to estimate quantitative trait locus parameters: power and genetic variances for unreplicated and replicated progeny. Genetics 126:769-777.

Knapp, S. J., W. C. Bridges and D. Birkes. 1990. Mapping quantitative trait loci using molecular marker linkage maps. Theor. Appl. Genet. 79: 583-592.

Knapp, S. J., W. C. Bridges and B. H. Liu. 1992. Mapping quantitative trait loci using nonsimultaneous and simultaneous estimators and hypothesis tests. In: J. S. Beckmann and T. C. Osborn (eds) Plant Genomes: Methods for Genetic and Physical Mapping. Kluwer

Academic Publishers, pp209-237.

Knott, S. A. 1994. Prediction of the power of detection of marker-quantitative trait locus linkage using analysis of variance. Theor. Appl. Genet. 89:318-322.

Knott, S. A. and C. S. Haley. 1992. Aspects of maximum likelihood methods for mapping of quantitative trait loci in line crosses. Genet. Res., Camb. 60:139-151.

Knott, S. A. and C. S. Haley. 1992. Maximum likelihood mapping of qunatitative traitloci using full-sib families. Genetics 132.1211-1222.

Knott, S. A. and C. S. Haley. 1992. Aspects of maximum likelihood interval mapping in an F_2 population. Genet. Res. 60:139-151.

Knutzon, D. S., G. A. Thompson, S.E. Radke, W.B. Johnson, V. C. Knauf and J.C. Kridl. 1992. Modification of *Brassica* seed oil by antisense expression ofa staryl-acyl carrier protein desaturase gene. Proc. Natl. Acad. Sci. USA 89:2624-2628.

Kosambi, D. D. 1944. The estimation of map distances from recombination values. Ann. Eugen. 12:172-175.

Kurata, N., G. Moore, Y. Nagamura, T. Foote, M. Yano, Y. Minobe and M. Gale. 1994. Conservation of genome structure between rice and wheat. Bio/technology 12:276-278.

Lamkey., K.R., A. R. Hallauer and A. L. Kahler. 1987. Allelic differences at enzyme loci and hybrid performance in maize. J. Heredity 78:231-234.

Lande, R. 1992. Marker-assisted selection in relation to traditional methods of plant breeding. In: H. T. Stalker and J. P. Murthy (eds) Plant Breeding in 1990s, CAB International Wallingford, UK, pp437-452.

Lander, E. S. and D. Botstein. 1986. Strategies for studying heterogeneous genetic traits in humans by using a linkage map of restriction fragment length polymorphisms. Proc. Natl. Acad. Sci. USA 83:7353-7357.

Lander, E. S. and D. Botstein. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185-199.

Lander, E. S. and P. Green. 1987. Construction of multilocus genetic linkage maps in humans. Proc. Natl. Acad. Sci. USA 84:2363-2367.

Lander, E. S., P. Green, J. Abrahamson, A. Barlow, M. J. Daly, S. E. Lincoln and L. Newburg. 1987. MAPMAKER : an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181.

Landry, B. S., R. V. Kesseli, B. Farrara and R. W. Michelmore. 1987. A genetic map of lettuce (*Lactuca sativa* L.) with restriction fragment length polymorphism, isozyme, disease resistance and morphological markers. Genetics 116:331-337.

Lathrop, G. M. and J. M. Labouel. 1984. Easy calculations of lod scores and genetic risks on small computers. Am. J. Hum. Genet. 36:460-465.

Law, C. N. 1966. The location of genetic factors affecting a quantitative character in wheat. Genetics 53:487-498.

Law, C. N. & M. D. Gale. 1979. Cytological markers and quantitative variation in wheat. In: J. N. Thompson and J. M. Thoday (eds) Quantitative Genetic Variation, Academic Press, New York. pp275-393.

Leary, J. J., D. J. Bright and D. C. Ward. 1983. Rapid and sensitive colorimetric method for

visualizing biotin-labelled DNA probes hybridized to DNA or RNA immobilized on nitrocellulose: bioblots. Proc. Natl. Acad. Sci. USA 80:4045-4049.

Lebowitz, R. L., M. Soller and J. S. Beckmann. 1987. Trait-based analysis for the detection of linkage between marker loci and quantitative trait loci in cross between inbred lines. Theor. Appl. Genet. 73:556-562.

Lerner, I. M. 1950. Population Genetics and Animal Improvement, Cambridge University Press, Cambridge.

Li, C. C. 1955. Population Genetics, University of Chicago Press, Chicago.

Liang, C. Z., M. H. Gu, X. B. Pan, G. H. Liang and L. H. Zhu, 1994. RFLP tagging of a new semidwarfing gene in rice. Theor. Appl. Genet. 88:898-900.

Lin, D. G. and S. Geng. 1986. Multivariate extension of Griffing's diallel analysis. J. Agron. Crop Sci. 157:52-57.

Lin C.S. and M. R. Binns. 1986. Stability analysis: where do we stand? Crop Sci. 26:894-900.

Lincoln, S. E., M. J. Daly and E. S. Lander. 1993. Mapping genes controlling quantitative traits using Mapmaker/QTL 1.1. Whitehead Institute Technical Report. 2nd edition.

Lloyd, A. M., V. Walbot and R. W. Davis. 1992. *Arabidopsis* and *Nicotiana* anthocyanin production activated by maize regulators *R* and *C1*. Science 258:1773-1775.

Lo Shiavo, F., G. Giulano and M. Terzi. 1983. Identifying natural and parasexual hybrids. In: S. D. Tanksley and T. J. Orton (eds) Isozymes in Plant Genetics and Breeding, Elsevier Science Publishers, Amsterdam, pp305-312.

Luo, Z. W. and M. J. Kearsey. 1989. Maximum likelihood estimation of linkage between a marker gene and a quantitative locus. Heredity 63:401-408.

Luo, Z. W. and M. J. Kearsey. 1991. Maximum likelihood estimation of linkage between a marker gene and a quantitative locus. II. Application to backcross and double haploid populations. Heredity 66:117-124.

Luo, Z. W. and M. J. Kearsey. 1992. Interval mapping of quantitative trait loci in an F_2 population. Heredity 69:236-242.

Luo, Z. W. and J. A. Woolliams. 1993. Estimation of genetic aprameters using linkage between a marker gene and a locus underlying a quantitative character in F_2 populations. Heredity 70: 245-253.

Lush, J. L. 1937. Animal Breeding Plans. Iowa State College Press. Ames, Iowa.

Lush, J. L. 1945. Animal Breeding Plans, 3rd ed., Iowa State College Press, Ames.

Mangin, B., B. Goffinet and A. Rebai. 1994. Constructing confidence intervals for QTL location. Genetics 138:1301-1308.

Manly, B. F. J. 1991. Randomization and Monte Carlo Methods in Biology. Chapman and Hall, London

Manly, K. F. 1993. A Macintosh program for storage and analysis of experimental genetic mapping data. Mamm. Genome 4:303-313.

Manly, K. F. 1994. Establishing genetic linkage using recombinant inbred lines with an abnormal segregation ratio. Genetics 136:1433-1434.

Mansur, L. M., J. Orf and K. G. Lark. 1993. Determining the linkage of quantitative trait loci to RFLP markers using extreme phenotypes of recombinant inbreds of soybean (*Glycine*

max L. Merr.). Theor. Appl. Genet. 86:914-918.

Markert, C. L. and F. Moller. 1959. Multiple forms of enzymes: tissue ontogenetic, and species specific patterns. Proc. Natl. Acad. Sci. USA 45:753-763.

Martin, B., J. Nienhuis, G. King and A.Schaefer.1989. Restriction fragmentlength polymorphisms associated with water use efficiency in tomato. Science 243:1725-1728.

Martin, G. B., J. G. K. Williams and S. D. Tanksley. 1991. Rapid identification of markers linked to a *Pseudomonas* resistance gene in tomato by using random primers and near-isogenic lines. Proc. Natl. Acad. Sci. USA 88:2336-2340.

Martin, G. B., S. H. Brommonschenkel, J. Chunwongse, A. Fray, M. W. Ganal, R. Spivey, T. Wu, E. D. Earle, S. D. Tamksley.1993. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262:1432-1436.

Martinez, O. and R. N. Curnow. 1992. Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor. Appl. Genet. 85: 480-488

Mather, K. 1949. Biometrical Genetics. Chapman & Hall, London.

Mather, K. and M. J. Hanks.1978. Genetics of coxal chaetae in *Drosophila melanogaster*: 1. Variation in gene action. Heredity 4:71-96.

Mather, K. and J. L. Jinks. 1971. Biometrical Genetics. Chapman & Hall,London.

Mather, K. and J. L. Jinks. 1982. Biometrical Genetics. Chapman & Hall, London.

McCouch, S. R., G. Kochert, Z. H. Yu., Z. Y. Wang, G. S. Khush, W. R. Coffman and S. D. Tanksley. 1988. Molecular mapping of rice chromosomes. Theor. Appl. Genet. 76: 148-149.

McMillan, I. and A. Robertson. 1974. The power of methods for the detection of major genes affecting quantitative traits. Heredity 32:349-356.

Melchinger, A. E. 1990. Use of molecular markers in breeding for oligogenic disease resistance. Plant Breed. 104:1-19.

Meng, X.-L. and D. B. Rubin. 1993. Maximum likelihood estimation via the ECM algorithm: a general framework. Biometrika 80:267-268.

Michelmore, R. W., I. Paran and R. V. Kesseli. 1991. Identification of markers linked to disease resistance gene by bulked segregant analysis : a rapid method to detect markers in specific genomic regions using segregating populations. Proc. Natl. Acad. Sci. USA 88: 9828-9832.

Mindrinos, M., F. Katagiri, G.-L. Yu and F. M. Ausubel. 1994. The *A. thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide-binding site and leucine-rich repeats. Cell 78:1089-1099.

Monaco, A. P. 1994. Isolation of genes from cloned DNA. Current Opinion in Genetics and Development 4:360-365.

Monaco, A. P. and Z. Larin. 1994. YACs, BACs, PACs and MACs: artificial chromosomes as research tools. TIBTECH 12:280-286.

Moore, G., M. D. Gale, N. Kurata and R. B. Flavell. 1993. Molecular analysis of small grain cereal genomes: current status and prospects. Bio/technology 11:584-589.

Moreno-Gonzalez, J. 1992. Estimates of marker-associated QTL effects in Monte Carlo backcross generations using multiple regression. Theor. Appl. Genet. 85:423-434.

Moreno-Gonzalez, J. 1993. Efficiency of generations for estimating marker-associated QTL

effects by multiple regression. Genetics 135:223-231.

Morton, N.E. 1955. Sequential test for the detection of linkage. Am.J. Hum. Genet. 7:277-318.

Motro, U. and M. Soller. 1993. Sequential sampling in determining linkage between marker loci and quantitative trait loci. Theor. Appl. Genet. 85:658-664.

Muehlbauer, G. J., J. E. Specht, M. A. Thomas-Compton, P. E. Staswick and R. L. Bernard. 1988. Near-isogenic lines - A potential resource in the integration of conventional and molecular marker linkage maps. Crop Sci. 28:729-735.

Murata, N. O. Ishizaki-Nishizawa, S. Higashi, H. Hayashi, Y. Tasaka and I.Nishida.1992. Genetically engineered alteration in the chilling sensitivity of plants. Nature 356:710-713.

Nakamura, Y., M. Leppert, P. O'Connell, R. Wolff, T. Holm, M. Culver, C. Martin, E. Fujimoto, M. Hoff, E. Kumlin and R. White.1987. Variable number of tandem repeat VNTR markers for human gene mapping. Science 235:1616-1622.

Nakamura, Y., M. Carlson, K.Krapcho, M. Kanamori and W.White. 1988. New approach for isolation of VNTR markers. Am. J. Hum. Genet. 43:854-859.

Nelson, R. R. 1979. The evolution of parasitic fitness. In: J. G. Horsfall and E. B. Cowling (eds) Plant Disease. An Advanced Treatise, Vol.4., Academic Press, New York, pp23-46.

Nienhuis, J., T. Helentjaris, M. Slocum, B. Ruggeroand A.Schaefer. 1987. Restriction fragment length polymorphism analysis of loci associated with insect resistance in tomato. Crop Sci. 27:797-803.

Nilsson-Ehle, H.1909.Inverstigations on crosses of oats and wheat, Lunds Univ. Arsskrift, n.s., ser.2, vol.5, no.2.

Olson, M., L. Hood, D. Cantor and D. Botstein. 1989. A common language for physical mapping of the human genome. Science 245:1434-1435.

Orita, M., H. Iwahana, H. Kanazawa, K. Hayashi and T. Sekiya, 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc. Natl. Acad. Sci. USA 86:2766-2770.

Orton, T. J. 1983. Applications of isozyme technology in breeding cross-pollinated crops. In: S. D. Tanksley and T. J. Orton (eds) Isozymes in Plant Genetics and Breeding, Elsevier Science Publishers, Amsterdam, pp363-380.

Osborn, T. C., D. C. Alexanderand J. F. Fobes. 1987. Identification of restriction fragment length polymorphisms linked to genes controlling soluble solids content in tomato fruit. Theor. Appl. Genet. 73:350-356.

Ott, J. 1985. Analysis of Human Genetic Linkage. The Johns Hopkins University Press.

Pachnis, V., L. Pevny, R. Rothstein and F. Constantni. 1990. Transfer of a yeast artificial chromosome carrying human DNA from *Saccharomyces cerevisiae* into mammalian cells. Proc. Natl. Acad. Sci. USA 87:5109-5113.

Parrish, J. E. and D. L. Nelson. 1994. Methods for finding genes: A major rate-limiting step in positional cloning. GATA 10(2):29-41.

Paterson, A. H., E. S. Lander, J. D. Hewitt, S. Peterson, S. E. Lincoln and S. D. Tanksley. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721-726.

Paterson, A. H., J. W. De Verna, B. Lanini and S. T. Tanksley. 1991. Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. Genetics 124: 735-742.

Paterson, A. H., S. Damon, J. D. Hewitt, D. Zamir, H. D. Rabinowitch, S. E. Lincoln, E. C. Lander and S. D. Tanksley. 1991. Mendelian factors underlying quantitative traits in tomato: comparison across species, generation, and environments. Genetics 127:181-197.

Peterson, P. A. 1992. Quantitative inheritance in the era of molecular biology. Maydica 37:7-18.

Plough, H. H. 1917. The effect of temperature on crossing over. J. Exp. Zool. 32:187-202.

Pollak, E., O. Kempthorne & T. B. Bailey, Jr. (ed.) 1977. Proceedings of the International Conference on Quantitative Genetics. Iowa State Univ. Press, Ames.

Powell, W., W. T. B. Thomas, D. M. Thompson, J. S. Swanston and R. Waugh. 1992. Association between rDNA alleles and quantitative traits in doubled haploid populations of barley. Genetics 130:187-194.

Ragot, M. and D. A. Hoisington. 1993. Molecular markers for plant breeding: comparisons of RFLP and RAPD genotyping costs. Theor. Appl. Genet. 86:975-984.

Rasmusson, J. M. 1935. Studies on the inheritance of quantitative characters in *Pisum*: I. Preliminary note on the genetics of flowering. Hereditas 20:161-180.

Rebai, A. and B. Goffinet. 1993. Power of tests for QTL detection using replicated progenies derived from a diallel cross. Theor. Appl. Genet. 86:1014-1022.

Rees, H. 1955. Genotypic control of chromosome behaviour in rye. I. Inbred lines. Heredity 9: 93-116.

Reiter, R. S., J. G. Coors, M. R. Sussman and W. H. Gabelman. 1991. Genetic analysis of tolerance to low-phosphorus stress in maize using restriction fragment length polymorphisms. Theor. Appl. Genet. 82:561-568.

Rendel, J. M. 1979. Canalization and selection. In: J. N. Thompson and J. M. Thoday (eds) Quantiative Genetic Variation, Academic Press, London, pp139-156.

Renz, M. and C. Kurz. 1984. A colorimetric method for DNA hybridization. Nucl. Acid Res. 12: 3435-3444.

Richards, E. J. and F. M. Ausubel. 1988. Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. Cell 53:127-136.

Rick, C. M. and J. F. Fobes. 1974. Association of an allozyme with nematode resistance. Rep. Tomato Genet. Coop. 24:25.

Ritland, K. 1983. Estimation of mating systems. In: S. D. Tanksley and T. J. Orton (eds) Isozymes in Plant Genetics and Breeding, Elsevier Science Publishers, Amsterdam, pp 289-302.

Ritter, E., C. Gebhardt and F. Salamini. 1990. Estimation of recombiantion frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. Genetics 125:645-654.

Robertson, D. S. 1985. A possible technique for isolating genic DNA for quantitative traits in plants. J. Theor. Biol. 117:1-10.

Robertson, D. S. 1989. Understanding the relationship between qualitative and quantitative

genetics. In: T. Helentjaris and B. Burr (eds) Development and Application of Molecular Markers to Problems in Plant Genetics, Cold Spring Harbor Laboratory, pp81-87.

Rodolphe, F. and M. Lefort. 1993. A multi-marker model for detectiong chromosome segments displaying QTL activity. Genetics 134:1277-1288.

Rommens, J. M., M. C.Iannuzzi, K. Bat-Sheva, M. L. Drumm, G. Melmer, M. Dean, R. Rozmahel, J. L. Cole, D. Kennedy, N. Hidaka, M. Zsiga, M. Buchwald, J. R. Riordan, L. C. Tsui and F. S Collins. 1989. Identification of the cystic fibrosis gene: chromosome walking and jumping. Science 245:1059-1065.

Saito, A., E. Shimosaka, Y. Hayano, K. Saito, S. Matuura and M. Yano.1991. Molecular mapping on chromosome 3 of rice. Jpn. J. Breed. 41:158-159 (Suppl 1).

Sax, K. 1923. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8:552-560.

Scaffer, J. F. and A.P. Roefls. 1985. Estimated relation between numbers of urediniospores of *Puccinia graminis* f. sp. *tritici* and rates of occurence of virulence. Phytopathology 75: 749-750.

Schmitz, S.-F. H., S.J. Schwager and E.J. Pollak. 1993.Determining the minimum sample size reqiured to obtain sufficient progeny with a derived genotype at two quantiative trait loci. Theor. Appl. Genet. 87:136-144.

Schwartz, D. C. and C. R. Cantor. 1984. Separation of yeast chromosome-sized DNAs by pulse field gradient gel electrophoresis. Cell 37:67-75.

Simmonds, N. W. 1979. Principles of Crop Improvment. Longman, London & New York.

Simpson, S. P. 1989. Detection of linkage between quantitative trait loci and restriction fragment length polymorphisms using inbred lines. Theor. Appl. Genet.77: 815-819.

Smith, J. S. C. and O. S. Smith. 1989. The use of morphological, biochemical,and genetic characteristics to measure distance and to test for minimum distance between inbred lines of maize (*Zea mays* L.). (Mimeo of paper presented at UPOV Workshop, Versailles, France, Octoer 1989). Pioneer Hi-Bred International. Inc., Johnston, IA.

Snape, J. W.1988. The detection and estimation of linkage using double haploid of single seed descent populations. Theor. Appl. Genet. 76:125-128.

Soller, M. and J. S. Beckmann. 1983.Genetic polymorphism in varietal identification and genetic improvement. Theor.Appl. Genet. 67:25-33.

Soller, M., T. Brody and A. Genizi. 1976. On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor. Appl. Genet. 47:35-39.

Sorrells, M. E. 1992. Development and application of RFLPs in polyploids. Crop Sci.32:1086-1091.

Stam, P. and A. C. Zeven. 1981. The theoretical proportion of the donor genome in near-isogenic lines of self-fertilizers bred by backcrossing. Euphytica 30:227-238.

Stark, D. M., K. P. Timmerman, G. F. Barry, J. Preiss and G. M. Kishore. 1992. Regulation of the amount of starch in plant tissues by ADP glucose pyrophosphorylase. Science 258:287-292.

Stuart, A. and J. K. Ord. 1991. Kendall's Advanced Theory of Statistics, 5th Ed., Vol.2.,

Oxford Univ. Press, New York.

Stuber, C. W. 1992. Biochemical and molecular markers in plant breeding. Plant Breed. Rev. 9: 37-61.

Stuber, C. W. and M. D. Edwards. 1986. Genotypic selection for improvement of quantitative traits in corn using molecular marker loci. Proc. 41st Annual Corn and Sorghum Research Conf., American Seed Trade Assoc. 41:70-83.

Stuber, C. W., R. H. Moll, M. M. Goodman, H. E. Schaffer, B. S. Weir. 1980. Allozyme frequency changes associated with selection for increased grain yield in maize (*Zea mays* L.). Genetics 95:225-236.

Stuber, C. W., S. E. Lincoln, D. W. Wolff, T. Helentjaris and E. C. Lander. 1992. Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823-839.

Sughroue, J. R. and T.R.Rocheford. 1994. Restriction fragment length polymorphism differences among Illinois Long-Term Selection oil strains. Theor. Appl. Genet. 87:916 -924.

Sun, T.-P., H. M. Goodman and F. M. Ausubel. 1992. Cloning the Arabidopsis *GAI* locus by genomic subtraction. Plant Cell 4:119-128.

Tanksley, S.D. 1983. Gene mapping. In: S. D. Tanksley and T. J. Orton (eds) Isozymes in Plant Genetics and Breeding, Part A. Elsevier Sci. Publishers. B. V. Amsterdam, pp109-139.

Tanksley, S. D. and J. Hewitt. 1988. Use of molecular markers in breeding for soluble solids content in tomato - a re-examination. Theor. Appl. Genet. 75:811-823.

Tanksley, S. D., H. Medina-Filho and C.M. Rick. 1982. Use of naturally-occurring enzyme variation to detect and map genes controlling quantitaitve traits in an interspecific backcross of tomato. Heredity 49:11-25.

Tanksley, S. D., J. Miller, A. Paterson and R. Bernatzky. 1988. Molecular mapping of plant chromosome. In: Gustafson, J. F. and R. Appels (eds.) Chromosome Structure and Function, Plenum, New York, pp157-173.

Tanksley, S. D., N. D. Young, A.H. Paterson and M. W.Bonierbale. 1989. RFLP mapping in plant breeding: Newtools for an old science. Bio/technology 7:257-263.

Tanksley, S. D., M. W. Ganal, J. P. Prince, M. C. de Vicente, M. W. Bonierbale, P. Broun, T.M. Fulton, J.J. Giovannoni, S. Grandillo, G. B. Martin, R. Messeguer, J. C. Miller, L. Miller, A. H. Paterson, P. Pineda, M. S. Roder, R. A. Wing, W. Wu and N. D. Young. 1992. High density molecular linkage maps of the tomato and potato genomes. Genetics 132:1141-1160.

Tautz, D. 1989. Hypervariability of simple sequences as a general source for polymorphic markers. Nucl. Acid Res. 17:6463-6471.

Taylor, B. A. 1978. Recombinant inbred strains: use in gene mapping. In: H. C. Morse (eds) Origins of Inbred Mice, Academic Press, New York. pp423-438.

Tease, C. 1978. Cytological detection of crossing-over in BudR substituted meiotic chromosomes using the fluorescent plus Giemsa technique. Nature 272:823-824.

Thoday, J. M. 1961. Location of polygenes. Nature 191:368-370.

Thoday, J. M. 1972. Disruptive selection. Proc. Roy. Soc., London, B. 182:109-143.

Thoday, J. M. 1979. Polygene mapping: uses and limitations In: J. N. Thompson and J. M.

Thoday (eds) Quantitative Genetic Variation, Academic Press, New York, pp219-233.

Van Ooijen, J. W. van. 1992. Accuracy of mapping quantitative trait loci in autogamous species. Theor. Appl. Genet. 84:803-811.

Wang, G.-L. and A. H. Paterson. 1994. Assessment of DNA pooling strategies for mapping of QTLs. Theor. Appl. Genet. 88:355-361.

Wang, G.-L., D. J. Mackill, J. M. Bonman, S. R. McCouch, M. C. Champoux and R. J. Nelson. 1994. RFLP mapping of genes conferring complete and partial resistance to blast in a durably resistant rice cultivar. Genetics 136:1421-1434.

Warner, J. N. 1952. A method for estimating heritability. Agron. J. 44:427-430.

Waterman, M. S. 1989. Mathematical Methods for DNA Sequences. CRC Press, Boca Raton, FL.

Waugh, R. and W. Powell. 1992. Using RAPD markers for crop improvement. TIBTECH 10:186-191.

Weber, J. L. 1990a. Human DNA polymorphisms based on length variations in simple-sequence tandem repeats. In: K. E. Davis and S. M. Tilghman (eds) Genome Analysis Vol.I. Genetic and Physical Mapping. Cold Spring Harbor Laboratory Press. pp159-181.

Weber, J. L. 1990b. Informativeness of human $(dC-dA)_n.(dG-dT)_n$ polymorphisms. Genomics 7: 524-530.

Weber, D. and T. Helentjrais. 1989. Mapping RFLP loci in mazie using B-A translocations. Genetics 121:583-590.

Weber, J. L. and P. E. May. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. 44:388-396.

Weeden, N. F. and R. Provvidenti. 1988. A marker locus, Adh-1, for resistance to peaenation mosaic virus in *Pisum sativum* . J. Heredity 79:128-131.

Wehrhahn, C. and R. W. Allard. 1965. The detection and measurement of the effects of individual genes involved in the inheritance of a quantitative character in wheat. Genetics 51:109-119.

Weide, R., M. F. Van Wordragen, R. K. Lankhorst, R. Verkerk, C. Hanhart, T. Liharska, E. Pap, P. Stam, P. Zabel and M. Koornneef. Integration of the classical and molecular linkage maps of tomato chromosome 6. Genetics 135:1175-1186.

Weir, B. S. 1990. Genetic Data Analysis, methodsfor discrete population genetic data, Sinauer Associates, Inc. Publishers, pp222-260.

Weir, B. S., E. J. Eisen, M. M. Goodman, G. Namkoong (eds). 1988. Proceedings of the Second International Conference on Quantitative Genetics. Sinauer, Sunderland, MA.

Weller, J. I. 1986. Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. Biometrics 42:627-641.

Weller, J. I. 1992. Statistical methodologies for mapping and analysis of quantitative trait loci. In: J. S. Beckmann and T. C. Osborn (eds) Plant Genomes: Methods for Genetic and Physical Mapping. Kluwer Academic Publishers, pp181-203.

Weller, J. I. and A. Wyler. 1992. Power of different sampling strategies to detect quantitative trait loci variance effects. Theor. Appl. Genet. 83:582-588,

Weller, J. I., Y. Kashi and M. Soller. 1990. Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. J. Dairy Sci. 73:2525-2537.

Welsh, J. and M. McClelland. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucl. Acid Res. 18:7213-7218.

Westcott, B. 1986. Some methods of analysing genotype-environment interactions. Heredity 56: 243-253.

Williams, J. G. K., A. R. Kubelik, K. J. Livak, J. A. Rafalski and S. V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl. Acid Res. 18:6531-6535.

Wilson, R., R. Ainscough, K. Anderson, C. Baynes, M. Berks, J. Bonfield, J. Burton, M. Connell, T. Copsey, J. Cooper, A. Coulson, M. Craxton, S. Dear, Z. Du, R. Durbin, A. Favello, A. Fraser, L. Fulton, A. Gardner, P. Green, T. Kawkins, L. Hillier, M. Jier, L. Johnston, M. Jones, J. Kershaw, J. Kirsten, N. Laisster, P. Latreille, J. Lightning, C. Lloyd, B. Mortimore, M. O'Callaghan, J. Parsons, C. Percy, L. Rifken, A. Roopra, D. Staden, J. Sulston, J. Thierry-Mieg, K. Thomas, M. Vaudin, K. Vaughan, R. Waterston, A. Watson, L. Weinstock, J. Wilkinson-Sproat and P. Wohldman. 1994. 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. Nature 368:32-38.

Wricke, G. and W. E. Weber, 1986. Quantitative Genetics and Selection in Plant Breeding. Walter de Gruyter, Berlin.

Wricke, G. and P. Wehling. 1985. Linkage between an incompatibility locus and a peroxidase isozyme locus (Prx7) in rye. Theor. Appl. Genet. 71:289-291.

Wright, A. J. and R. P. Mowers. 1994. Multiple regression for molecular-marker, quantitative trait data from large F_2 populations. Theor. Appl. Genet. 89:305-312.

Wu, K. S. and S. D. Tanksley. 1993. Genetic and physical mapping of telemeres and macrosatellite of rice. Plant Mol. Biol. 22:861-872.

Wu, K. S., J. C. Glaszmann and G. S. Khush. 1988. Chromosomal localization of ten isozyme loci in rice (*Oryza sativa* L.) through trisomic analysis. Biochem. Genet. 26:303-320.

Wu, W. R. and W. M. Li. 1994. A new approach for mapping quantitative trait loci using complete genetic marker linkage maps. Theor. Appl. Genet. 89:535-539.

Xu, Jichen, Lihuang Zhu, Ying Chen, Hongwei Cai, Yunbi Xu, Chaofu Lu, Ming Yang and Zhongzhuan Ling. 1994. Construction of a rice molecular linkage map using double haploid population. Chinese J. Genet. 21:127-136.

Xu, Yunbi. 1989. Allelism test for polygenes and its application to plant breeding, In : Proceedings of the 6th International Congress of SABRAO, P693-696.

Xu, Y. B. and Z. T. Shen. 1991. Diallel analysis of tiller number at different growth stages in rice (*Oryza sativa* L.). Theor. Appl. Genet. 83:243-249.

Xu, Yunbi and Shen Zongtan 1992a. Detection and genetic analyses of the gene dispersed cross: some theoretical considerations. Acta Agriculturae Universitatis Zhejiangensis, 18:109-117.

Xu Yunbi and Shen Zongtan 1992b. Detection and genetic analysis of the gene dispersed crosses for tiller angle in rice (*Oryza sativa* L.), Acta Agriculturae Zhejiangensis 4:54-60.

Xu Yunbi and Shen Zongtan 1992c. Accumulation of the alleles with similar effects at four loci controlling tiller angle from gene dispersed crosses in rice (*Oryza sativa* L.), J.

Biomath. (Beijing) 7(4):1-10.

Xu, Yun-Bi, Zong-Tan Shen, Ji-Chen Xu, Ying Chen and Li-Huang Zhu. 1993. Mapping quantitative trait loci via restriction fragment length polymorphism markers in rice. Rice Genet. Newsl. 10:135-138.

Young, N.D. and S. D. Tanksley. 1989a. Restriction fragment length polymorphism maps and the concept of graphical genotypes. Theor. Appl. Genet. 77:95-101.

Young, N.D. and S. D. Tanksley. 1989b. RFLP analysis of the size of chromosomal segments retained around the *Tm-2* locus of tomato during backcross breeding. Theor. Appl. Genet. 77:353-359.

Young, N.D., J. C. Miller and S. D. Tanksley. 1987. Rapid chromosomal assignment of multiple genomic clones in tomato using primary trisomics. Nucl. Acids Res. 15:9339- 9348.

Young, N.D., D. Zamir, M. Ganal and S. D.Tanksley. 1988. Use of isogenic lines and simultaneous probing to indentify DNA markers tightly linked to the *Tm-2a* gene in tomato. Genetics 120:579-585.

Zamir, D. 1983. Pollen gene expression and selection: applications in plant breeding. In: S. D. Tanksley and T. J.Orton (eds) *Isozymes* in Plant Genetics and Breeding, Elsevier Science Publishers, Amsterdam, pp313-330.

Zeng, Z.-B. 1993. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc. Natl. Acad. Sci. USA 90:10972-10976.

Zeng, Z.-B. 1994. Precision mapping of quantitative trait loci. Genetics 136:1457-1468.

Zeven, A. C., D. R. Knott and R. Johnson. 1983. Investigation of linkage drag in near *isogenic* lines of wheat by testing for seedling reaction to races of stem rust, leaf rust and yellow *rust*. Euphytica 32:319-327.

Zhang, Q., M. A. Saghai Maroof, T. Y. Lu and B. Z. Shen.1992.Genetic diversity and differentiation of *indica* and *japonica* rice detected by RFLP analysis. Theor. Appl. Genet. 83:495-499.

Zhang, Q., M. A. Saghai Maroofand A. Kleinhofs. 1993. Comaprative diversity analysis of RFLPs and isozymes within and among populations of *Hordeum vulgare* ssp. *spontaneum*. Genetics 134:909-916.

Zhang, Q., Y. J. Gao, S. H. Yang, R. A. Ragab, M. A. Saghai Maroof and Z. B. Li. 1994a.A diallel analysis of heterosis in elite hybrid rice based on RFLPs and microsatellites. Theor. Appl. Genet. 89:185-192.

Zhang, Q., B. Z. Shen, X. K. Dai, M. H. Mei, M. A. Saghai Maroofand Z. B. Li. 1994b.Using bulked extremes and recessive class to map genes for photoperiod-sensitive genic male sterility in rice. Proc. Natl. Acad. Sci. USA 91:8675-8679.

Zhang, W. and C. Smith. 1993. Simulation of marker-assisted selection utilizing linkage disquilibrium: the effects of several additional factors. Theor. Appl. Genet. 86:492-496.

Zhang, X. F., J. A. Mosjidis and Z. L. Hu. 1992. Methods for detection and estimation of linakge betweena markerlocus and quantitative trait loci. Plant Breed. 109:35-39.

Zheng, K., H. Qian, B. Shen and J. Zhuang. 1994. RFLP-basedphylogenetic analysis of wide compatibility varieties in *Oryza sativa* L. Theor. Appl. Genet. 88:65-69.

Zhu, Li-Huang, Ying Chen, Yun-Bi Xu, Ji-Chen Xu, Hong-Wei Cai and Zhong-Zhuan Ling.1993.

Construtionof a molecular map of rice and gene mapping using a double haploid population of a cross between Indica and Japonica varieties. Rice Genet. Newsl. 10: 132-135.

Zhuchenko, A. A., A. B. Korol and V. K. Andryushchenko.1978. Linkage between loci of quantitative characters and markerloci. Genetika 14:771-778.